CONTENTS OF VOLUME 82

Adams, J. A., McIntyre, J. S., and Thorsheim, H. I. Response Feedback and Verbal Retention 290
Ahmed, A., and Voss, J. F. Serial Acquisition as a Function of Stage of Learning 398
Amsel, A., Surridge, C. T., and Hug, J. J. Number of Food Pellets and the Partial Reinforcement Extinction Effect after Extended Acquisition 578
Anderson, N. H. See Weiss, D. J.
Angel, R. W., and Higgins, J. R. Correction of False Moves in Pursuit Tracking 185
Banks, W. P. Criterion Change and Response Competition in Unlearning 216
Baron, M. R. See Rittle, R. H.
Barrett, H. R. See Pitz, G. F.
Batchelder, W. H. See Nelson, T. O.
Baumeister, A. A., and Joubert, C. E. Interactive Effects on Reaction Time of Preparatory Interval Length and Preparatory Interval Frequency 393
Blick, K. I. Decision and Decay Processes in the Short-Term Memory for Length 224
Bolton, L. S. See Bower, G. H.
Bourne, L. E., Jr., Ekstrand, B. R., and Montgomery, B. Concept Learning as a Function of the Conceptual Rule and the Availability of Positive and Negative Instances 538
Bower, G. H., and Bolton, L. S. Why are Rhymes Easy to Learn? 453
Breen, W. W., De Haemer, M. J., and Poock, G. K. Comparison of the Effect of Auditory versus Visual Stimulation on Information Capacity of Discrete Motor Responses 395
Burrows, D., and Murdock, B. B., Jr. Effects of Extended Practice on High-Speed Scanning 231
Campbell, B. A., and Messing, R. B. Aversion Thresholds and Aversion Difference Limens for White Light in Albino and Hooded Rats 353
Campione, J. C., and Wentworth, P. C. Differential Cue Habit Strength as a Determinant of Attention 527
Card, S. K. See Clark, H. H.
Carvellas, T. See Kaplan, I. T.
Cermak, L. S. Repetition and Encoding in Short-Term Memory 321
Clark, H. H. Influence of Language on Solving Three-Term Series Problems 205
Clark, H. H., and Card, S. K. Role of Semantics in Remembering Comparative Sentences 545
Clark, S. E. Retrieval of Color Information from Preperceptual Memory 263
Cohen, G. Pattern Recognition: Differences between Matching Patterns to Patterns and Matching Descriptions to Patterns 427
Connor, W. H., and Lang, P. J. Cortical Slow-Wave and Cardiac Rate Responses in Stimulus Orientation and Reaction Time Conditions 310
Cramer, P. Functioning of Implicit Associative Responses in Mediated Transfer 301
Crothers, E. J. Arbitrary CVC Hierarchies Learned by Paired-Component Presentations 285
Crowder, R. G. Improved Recall for Digits with Delayed Recall Cues 258
Dammann, C. A., and Perkins, C. C., Jr. Duration of Antecedent Discriminative Stimuli and within-Subject Reward Magnitude Differences as Determiners of Running Speed .. 554
Dawson, M. E., and Reardon, P. Effects of Facilitory and Inhibitory Sets on GSR Conditioning and Extinction 462
De Haemer, M. J. See Breen, W. W.
Deregowski, J. B. Perception of the Two-Pronged Trident by Two- and Three-Dimensional Perceivers 9
DeRosa, D. V. Transformations on Sets in Short-Term Memory: Set Size Reduction by Deletion 415
Dick, A. O. Relations between the Sensory Register and Short-Term Storage in Tachistoscopic Recognition 279
Di Vesta, F. J., and Ingersoll, G. M. Effect of Semantic Redundancy on Children's Identification of Verbal Concepts 360
Dixon, T. R. See Neman, R.
Doll, T. J. Short-Term Retention: Preparatory Set as Covert Rehearsal 175
Dominowski, R. L. Concept Attainment as a Function of Instance Contiguity and Number of Irrelevant Dimensions 573
Donaldson, W., and Glathe, H. Recognition Memory for Item and Order Information 557
Eck, K. O., Noel, R. C., and Thomas, D. R. Discrimination Learning as a Function of Prior Discrimination and Nondifferential Training 156
Eimas, P. D. Observing Responses, Attention, and the Overtraining Reversal Effect 499
Ekstrand, B. R. See Bourne, L. E., Jr.
Epstein, W. Poststimulus Output Specification and Differential Retrieval from Short-Term Memory 168
Fell, J. C. See Laughery, K. R.
Filion, R. D. L., Fowler, S. C., and Notterman, J. M. Some Effects of Simultaneous Force-Proportional Positive and Negative Reinforcement 267
Fillenbaum, S. Words as Feature Complexes: False Recognition of Antonyms and Synonyms 400
Fisher, B. E. See Fisher, G. L.
Fisher, G. L., and Fisher, B. E. Differential Rates of GSR Habituation to Pleasant and Unpleasant Sapid Stimuli 339
Fowler, S. C. See Filion, R. D. L.
Frankel, A. S. Effects of Arousal on Hierarchically Organized Responses 385
Friedrichs, A. G. See Zeiler, M. D.
Fuchs, A. H. Recall for Order and Content of Serial Word Lists in Short-Term Memory .. 14
Garland, R. M. See Nelson, D. L.

Gholson, B., and Hohle, R. H. Choice Response Times as Functions of Intralist Similarity, Stimulus Type, and Number of Equally Probable Alternatives 581
Glathe, H. See Donaldson, W.
Grant, D. A., Kroll, N. E. A., Kantowitz, B., Zajano, M. J., and Solberg, K. B. Transfer of Eyelid Conditioning from Instrumental to Classical Reinforcement and Vice Versa 503
Grasha, A. F. See Schumsky, D. A.
Gregg, R. E. See Murray, F. S.
Gumenik, W. E. Effects of Articulatory Activity and Auditory, Visual, and Semantic Similarity on the Short-Term Memory of Visually Presented Paired Associates 70
Guy, D. E. Developmental Study of Performance on Conceptual Problems involving a Rule Shift 242
Hamilton, P. Selective Attention in Multisource Monitoring Tasks 34
Harding, G. B., and Rundle, G. R. Long-Term Retention of Modality- and Nonmodality-Specific Habituation of the GSR 390
Hawkins, H. L., and Hosking, K. Stimulus Probability as a Determinant of Discrete Choice Reaction Time 435
Headley, D. B. See Stanners, R. F.
Henion, K. E. Subjective Range of the Odor Oiliness of Heptanol 515
Henschel, D. See Postman, L.
Higgins, J. R. See Angel, R. W.
Hoats, D. L. See Thor, D. H.
Hohle, R. H. See Gholson, B.
Hosking, K. See Hawkins, H. L.
Houston, B. K. Noise, Task Difficulty, and Stroop Color-Word Performance 403
Houston, J. P. Proactive Inhibition and Undetected Retention Interval Rehearsal 515
Huang, I.-N. Successive Contrast Effects as a Function of Type and Magnitude of Reward 64
Hudson, R. L. Category Clustering for Immediate and Delayed Recall as a Function of Recall Cue Information and Response Dominance Variability 575
Hug, J. J. See Amsel, A.
Hyde, T. S., and Jenkins, J. J. Differential Effects of Incidental Tasks on the Organization of Recall of a List of Highly Associated Words 472
Ingersoll, G. M. See Di Vesta, F. J.
Jenkins, J. J. See Hyde, T. S.
Joubert, C. E. See Baumeister, A. A.
Kantowitz, B. See Grant, D. A.
Kantowitz, B. H. Double Stimulation with Varying Response Information 347
Kaplan, I. T., Carvellas, T., and Metlay, W. Searching for Words in Letter Sets of Varying Size 377
Karabenick, S. A. Effects of Reward Increase and Reduction in the Double Runway 79
Kausler, D. H., and Olson, R. D. Homonyms as Items in Verbal Discrimination Learning and Transfer 136
Keeney, T. J. Permutation Transformations on Phrase Structures in Letter Sequences 28
Keppel, G., and Zavortink, B. Further Test of the Use of Images as Mediators 190
Klipple, A. G., and Roberts, K. M. Effects of Event Probability and Cost on Performance in a Continuous Motor Task 75
Kohfeld, D. L. Effects of the Intensity of Auditory and Visual Ready Signals on Simple Reaction Time 88
Kroll, N. E. A. See Grant, D. A.
Lang, P. J. See Connor, W. H.
Laughery, K. R., and Fell, J. C. Subject Preferences and the Nature of Information Stored in Short-Term Memory 193
Levine, M. Latency-Choice Discrepancy in Concept Learning 1
Liu, I.-M. Cueing Function of Fragments of Verbal Items 107
Mackavey, W. R. Spacial Brightness Changes in Koffka's Ring 405
Mann, R. See Platt, J. R.
Martin, D. W. Data Conflict in a Multinomial Decision Task 4
McConkie, G. W. Effects That a Massed Repetition of One Pair Has on Other Pairs in a List 187
McIntyre, J. S. See Adams, J. A.
Messing, R. B. See Campbell, B. A.
Metlay, W. See Kaplan, I. T.
Meunier, G. F. See Stanners, R. F.
Modigliani, V., and Saltz, E. Evaluation of a Model Relating Thorndike-Lorge Frequency and *m* to Learning 570
Montgomery, B. See Bourne, L. E., Jr.
Murch, G. M. Growth of a Percept as a Function of Interstimulus Interval 121
Murdock, B. B., Jr. See Burrows, D.
Murray, F. S., and Gregg, R. E. Reception versus Selection Procedures in Concept Learning 571
Nelson, D. L., and Garland, R. M. Amount and Locus of Stimulus–Response Overlap in Paired-Associate Acquisition 297
Nelson, T. O. Dimensional Similarity in Concept Identification and Extradimensional Shifts 143
Nelson, T. O., and Batchelder, W. H. Forgetting in Short-Term Recall: All-or-None Decremental? 96
Neman, R., and Dixon, T. R. Idiodynamic and Popular Sets in Word Associations of College Students 381
Noel, R. C. See Eck, K. O.
Notterman, J. M. See Filion, R. D. L.
Olson, R. D. See Kausler, D. H.
Paivio, A. See Yuille, J. C.
Perkins, C. C., Jr. See Dammann, C. A.

Pishkin, V., and Shurley, J. T. Hydro-Hypodynamic Sensory Isolation Effects on Concept Identification ... 198
Pitz, G. F., and Barrett, H. R. Information Purchase in a Decision Task following the Presentation of Free Information ... 410
Platt, J. R., Senkowski, P. D., and Mann, R. Discrimination, Contrast, and Chaining: Effects of Prior Training without Discriminanda and Response-Contingent Delay of Discriminandum Presentation ... 38
Poock, G. K. See Breen, W. W.
Postman, L. See Monographs listing.
Reardon, P. See Dawson, M. E.
Remington, R. J. Analysis of Sequential Effects in Choice Reaction Times ... 250
Richman, C. L. See Schumsky, D. A.
Rickert, E. J. Role of Feedback Stimuli in Response Discrimination and Differentiation .. 148
Rittle, R. H., and Baron, M. R. Selective Attention and Rate of Discrimination Learning as a Function of Intradimensional Variability ... 327
Robbins, J. Probability Learning in the Correction T Maze under Noncontingent Reinforcement Schedules ... 115
Roberts, K. M. See Klipple, A. G.
Ross, L. E. See Wilcox, S. M.
Rundle, G. R. See Harding, G. B.
Runquist, W. N. Positive and Negative Transfer Resulting from Formal Similarity of Stimuli ... 129
Saltz, E. See Modigliani, V.
Schaeffer, B., and Wallace, R. Semantic Similarity and the Comparison of Word Meanings 343
Schumsky, D. A., Grasha, A. F., Trinder, J., and Richman, C. L. List Length and Single-Trial Short-Term Memory ... 238
Senkowski, P. D. See Platt, J. R.
Shepp, B. E., and Turrisi, F. D. Effects of Overtraining on the Acquisition of Intradimensional and Extradimensional Shifts ... 46
Shurley, J. T. See Pishkin, V.
Smith, M. C. Effect of Varying Channel Capacity on Stimulus Detection and Discrimination 520
Solberg, K. B. See Grant, D. A.
Spitz, H. H. Effects of Stimulus Information Reduction on the Search Time of Retarded Adolescents and Normal Children ... 482
Stanners, R. F., Meunier, G. F., and Headley, D. B. Reaction Time as an Index of Rehearsal in Short-Term Memory ... 566
Stark, K. See Postman, L.
Steinbach, M. J. Eye Tracking of Self-Moved Targets ... 366
Strongman, K. T., and Wookey, P. E. Frustrative Reward ... 183
Surridge, C. T. See Amsel, A.
Thomas, D. R. See Eck, K. O.
Thomas, H. Preference following Adaptation of an Ideal and the Metric Scaling of Visual Complexity ... 488
Thor, D. H., Winters, J. J., Jr., and Hoats, D. L. Vertical Eye Movement and Space Perception ... 163
Thorsheim, H. I. See Adams, J. A.
Trinder, J. See Schumsky, D. A.
Turrisi, F. D. See Shepp, B. E.
Vandament, W. E. Response Latency as a Function of Interstimulus Interval in Conditioned Eyelid Discrimination ... 561
Voss, J. F. See Ahmed, A.
Wallace, R. See Schaeffer, B.
Weisberg, R. W. Sentence Processing Assessed through Intrasentence Word Associations 332
Weiss, D. J., and Anderson, N. H. Subjective Averaging of Length with Serial Presentation 52
Weiss, S. J. Attentional Processes along a Composite Stimulus Continuum during Free Operant Summation ... 22
Wentworth, C. See Campione, J. C.
Wilcox, S. M., and Ross, L. E. Differential Classical Eyelid Conditioning as a Function of CS Intensity, CS Rise Time, and Interstimulus Interval ... 272
Winters, J. J., Jr. See Thor, D. H.
Wood, G. Whole–Part Transfer from Paired-Associate to Free Recall Learning ... 532
Wookey, P. E. See Strongman, K. T.
Yuille, J. C., and Paivio, A. Abstractness and the Recall of Connected Discourse ... 467
Zacks, R. T. Invariance of Total Learning Time under Different Conditions of Practice ... 441
Zajano, M. J. See Grant, D. A.
Zavortink, B. See Keppel, G.
Zeiler, M. D., and Friedrichs, A. G. Absolute and Relational Discriminations involving Three Stimuli ... 448
Zippel, B. Unrestricted Classification Behavior and Learning of Imposed Classifications in Closed, Exhaustive Stimulus Sets ... 493

Monographs

Postman, L., Stark, K., and Henschel, D. Conditions of Recovery after Unlearning. No. 1, Part 2.

Journal of
Experimental Psychology

VOL. 82, NO. 1, PART 1 OCTOBER 1969

333
24.8.
Lelo

LATENCY-CHOICE DISCREPANCY IN CONCEPT LEARNING[1]

MARVIN LEVINE[2]

State University of New York at Stony Brook

During concept learning, the probability of the correct response shifts abruptly from .5 to 1.0 around the trial of last error. This shift has been effectively treated by a theory of discontinuous processes. Latency measures, however, show a gradual decrease during the criterion run of correct responding. The gradual decrease can be explained by assuming that *S* is monitoring a set of hypotheses (Hs) and is narrowing the set down until only the correct H remains. An implication of this view, that the latencies will be constant from the point at which *S* holds only the correct H, was verified.

In recent years, evidence has accrued for the view that adult humans learn concepts by a discontinuous process (specifically, by hypothesis testing). This formulation received strong support from the demonstration by Bower and Trabasso (1964) that the trial of the last error (TLE) in a protocol was an important pivot around which the discontinuity appeared. They showed the proportion of correct responses at each trial before the TLE to be constant at about .5. After the TLE the proportion is 1.0, by definition. Thus, plotting around the TLE in this way yields a two-part curve. Bower and Trabasso accounted for this curve and for other data by assuming that *S* sampled an hypothesis (H) from a set of Hs and that the H dictated *S*'s response. If the response was correct, *S* kept his H; if it was wrong, *S* returned the H to the set and randomly resampled for a new one. Incorrect Hs led to an error with $P = .5$. Only the correct H permitted *S* to have a long unbroken string of correct responses. The obtained discontinuity around the TLE, then, was paralleled in the theory: *S* goes from holding incorrect Hs to holding the correct H. For present considerations, the important feature of this theory is that *S* holds one H at a time.

Latencies should also reflect this discontinuity. As Erickson, Zajkowski, and Ehmann (1966) pointed out, latencies after the TLE should be constant and should be lower than the average latency before the TLE. This result, however, does not obtain. Erickson et al. showed that the latency during the criterion run, when the *S* is presumed to be holding the single correct H, is gradually decreasing over trials. Thus, a puzzle was posed. Choice responses reflect a discontinuous process; latencies show decreases that suggest gradual strengthening.

The discrepancy is resolved by abandoning the "one H at a time" conception. An earlier version of H theory (Restle, 1962) had included the possibility that *S* might

---

[1] This research was supported by United States Public Health Service Grant MH-11857-02. Gratitude is owed Noel Leffler, who supervised the experiment.

[2] Requests for reprints should be sent to Marvin Levine, Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11790.

© 1969 by the American Psychological Association, Inc.

select and monitor a small subset of Hs. Levine (1966) and Trabasso and Bower (1968) showed that the data require this assumption. Levine (in press) has postulated that *S* (*a*) samples a subset of the Hs, (*b*) takes one, a working H, from this subset as the basis for his response, and (*c*) uses the outcome, whether "right" or "wrong," to evaluate the Hs in the subset. Those Hs disconfirmed by the outcome may be rejected, thereby reducing the subset. Whenever a response is called "wrong," the working H is among those disconfirmed, so that a new one is taken. According to this view, the TLE is that trial at which *S* first takes the correct H as his new working H. During the next few trials, he is reducing his subset until none but the working H remains. Trabasso and Bower made the reasonable assumption that the latency is a function of the number of Hs in the subset to be evaluated. This implies that there should be a decrease in latency over the first few trials of the criterion run.

The subset assumption, then, nicely accounts for the decreasing latency curve. The theory, however, has another implication. If the TLE is that trial on which *S* first takes the correct H as his working H, then there is a later trial at which all other Hs are eliminated. At that trial, the solution trial (ST), *S* first holds a single H. Thus, the ST, like the TLE, is also a pivotal trial. Before it, *S* has a decreasing subset containing more than the working H; after it, *S* holds only the working H. The latencies should reflect this two-state character by a two-part curve: From the TLE to the ST the latencies should decrease; beyond the ST the latencies should be constant. Because the ST occurs at different trials for different *S*s, averaging at each trial after the TLE will obscure the discontinuity and yield a decaylike curve. However, averaging separately before and after the ST should yield the two parts clearly.

The two-part implication could be tested only if the ST could be located. A simple expedient was tried. A bell was placed near *S* engaged in a concept learning problem. He was told to ring the bell when he had the solution. If the trial on which *S* sounds the bell is indeed the ST and if the foregoing theory is correct, then averaging the latencies at each trial around the ST should yield the two-part curve.

METHOD

Subjects.—Thirty-two undergraduates volunteered for the experiment to fulfill a requirement for the introductory psychology course.

Materials.—The stimulus materials were on slides and were rear-projected onto two 3-in.-square translucent windows, set 6 in. apart. The display consisted of two letters (A or U), one large (1 in. high) and one small ($\frac{1}{2}$ in. high), one solid and one dashed, with one appearing in the left window and the other in the right. The attributes and letters were shifted from trial to trial in a quasi-random way, so that each attribute appeared with each letter and each position two times in every four trials. Enough stimulus slides were constructed to produce 10 32-trial problems.

Procedure.—Each *S* was seated before the panel containing the two windows. A sample slide was shown to *S*, and the eight characteristics (two positions, letters, sizes, and textures) were pointed out. He was instructed to press one of the two windows and that the word "right" or "wrong" would then light up, indicating whether he had made the correct response. He was further informed that one of the eight characteristics would be the basis for choosing correctly throughout a problem and that his task was to find the solution and to choose correctly all the time.

Sixteen *S*s (the E group) then had their attention drawn to a button located to their right and were told, "As soon as you know the solution, press this button. After pressing, tell me what the solution is, then continue with the problem." For the remaining 16 *S*s (the C group), this instruction was omitted. All *S*s then received two practice problems. After *S* in the E group rang the bell and announced the solution, *E* said simply, "Continue please." The *E* stated the solution ("A" and "right side") after each of these two preliminary problems. The *S* was then reminded of the eight possible solutions, and eight experimental problems were presented. A problem lasted either until *S* made 12 consecutive correct responses or for 32 trials. Each of the eight solutions was used once in each problem. Different *S*s had different sequences of solutions over the eight problems.

RESULTS AND DISCUSSION

The bell ringing did not appear to disrupt the performance of the E-group *S*s. They learned during the two practice problems

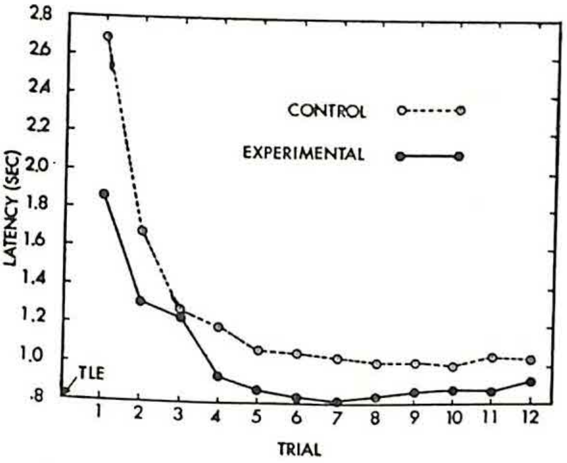


FIG. 1. Latencies during the criterion trials for the two groups.

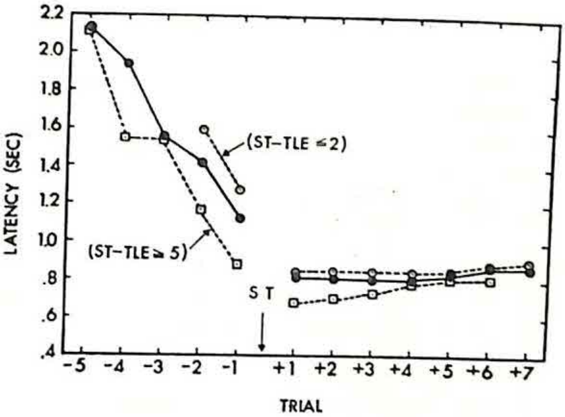


FIG. 2. Solid curve: latencies for the E group averaged at successive trials before and after the ST. Dashed curves: comparable data for two subgroups differing in the number of trials from the TLE to the ST.

to continue with the correct choices after announcing the solution. Only once did an *S* fail to ring the bell during the criterion run; only once did an *S* state the wrong solution. The influence of the bell procedure may be further evaluated from Fig. 1, which shows the post-TLE latencies for both groups. Except that *S*s in the E group are responding somewhat faster than *S*s in the C group, the two curves show similar curvature and both replicate the type of curve found by Erickson et al. (1966).

Figure 1 was constructed by aligning all the protocols at the TLE and averaging the latencies trial by trial (omitting in the E group the STs). The two-part hypothesis, however, may be tested by aligning the protocols at the ST and then averaging. This result is shown by the solid curve in Fig. 2. The two-part curve confirms the hypothesis with no apparent ambiguity. The possibility remains, however, that the decreasing solid line to the left of the ST could to be an artifact of averaging. To check on this possibility, the problems were separated into those in which *S* rang the bell shortly (0–2 trials; $n = 41$) after the TLE and those in which he rang it several trials (5 or more; $n = 15$) after the TLE. The results, presented as the dotted lines in Fig. 2, indicate that the ST is indeed a turning point in the criterion run, no matter where the ST occurs.

The results thus confirm the conception that *S* monitors a subset of Hs from which he selects a working H. The information at the end of each trial permits him to discard disconfirmed Hs from the subset. Latency is a function of the number of Hs to be evaluated in the subset. On this view, the reduction in latency reflects the reduction of the H set until a single H remains. It is only after this point that the condition envisaged by Erickson et al. (1966) appears: the latencies are constant.

REFERENCES

BOWER, G., & TRABASSO, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford, Calif.: Stanford University Press, 1964.

ERICKSON, J. R., ZAJKOWSKI, M. M., & EHMANN, E. D. All-or-none assumptions in concept identification. *Journal of Experimental Psychology,* 1966, 72, 690–697.

LEVINE, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology,* 1966, **71**, 331–338.

LEVINE, M. Neo-noncontinuity theory. In G. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 3. New York: Academic Press, 1969, in press.

RESTLE, F. The selection of strategies in cue learning. *Psychological Review,* 1962, **69**, 329–343.

TRABASSO, T., & BOWER, G. H. *Attention in learning.* New York: Wiley, 1968.

(Received January 16, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 4–8

DATA CONFLICT IN A MULTINOMIAL DECISION TASK[1]

DAVID W. MARTIN[2]

Ohio State University

Nine experienced and nine naive *S*s gave subjective posterior probability estimates to six hypotheses based on six items of evidence. Five levels of data conflict and three levels of sample diagnosticity were combined factorially with 25 trials per cell. A logarithmic payoff was used. When compared to Bayes' model, experienced *S*s became increasingly conservative and endorsed fewer correct hypotheses as conflict and sample diagnosticity increased. Naive *S*s also responded conservatively, but remained stable across conflict levels. The results are interpreted in terms of specific negative transfer of training and an inappropriate weighting of item impact.

When man is given evidence and required to estimate the likelihood of each event from a mutually exclusive and exhaustive set of events, he has been found to be conservative with respect to an optimal model (Bayes' theorem, Peterson & Beach, 1967). This model written in terms of the posterior odds (Ω_1) of two events or hypotheses given a sample of several items of evidence or data is

$$\Omega_1 = L\,\Omega_0,$$

where L is the likelihood of the sample of data given the occurrence of each of the hypotheses and Ω_0 is the odds of the two hypotheses prior to information concerning the data.

Consider, e.g., two urns (hypotheses) each containing 100 objects differing in some dimension such as color. Urn A might contain 60 red and 40 blue objects, while Urn B might contain 40 red and 60 blue objects. An urn is chosen based on the toss of a fair coin, and an object is drawn from the chosen urn. An *S* is told the color of the object and asked how likely it is that each urn was chosen given this information. In this example $\Omega_0 = 1/1$ and $L = l = 60/40$ or $40/60$. If several objects are drawn with replacement, L can be found by multiplying each of the individual likelihood ratios, e.g., the likelihood of the predominately red urn given two red objects and a blue, $L = \pi_i l_i = (60/40)(60/40)(40/60)$.

Thus, Ω_1 is determined by Ω_0 and L, where L is made up of the likelihood ratio of each datum (l), the number of data, and the manner in which each of these data favors a given hypothesis.

A number of past experiments have investigated three of the previous components that make up Ω_1. Generally, as Ω_0 departs from 1.0, conservatism increases (Peterson & Miller, 1965), although with a more restricted range, Ω_0 seems to have little effect (Phillips & Edwards, 1966; Schum, 1966). Perhaps the most pervasive effect on *S*'s performance has been found to be the size of l. As l approaches 1.0, *S*s become less conservative and even excessive in their estimates (Peterson & Miller, 1965; Phillips & Edwards, 1966; Schum & Martin, 1968). The number of data, or sample size, also affects conservatism. Peterson, Schneider, and Miller (1965) and Schum (1966) in widely differing contexts reported increased conservatism with increasing sample size.

The final component determining Ω_1 is the manner in which each l favors a given hypothesis. This component is called conflict and is defined as the number of l favoring any hypothesis other than the one generating the data sample; e.g., (60/40)(60/40)(60/

---

[1] This research was carried out at the Human Performance Center and was supported in part by the Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, under Contract AF 33(615)-2248 with the Ohio State University Research Foundation. Further reproduction is authorized to satisfy needs of the United States Government. An expanded version of this study was submitted as partial fulfillment of the requirements for the MA degree.

[2] Requests for reprints should be sent to David W. Martin, Human Performance Center, Ohio State University, 404-B West 17th Avenue, Columbus, Ohio 43210.

40) has less conflict than (60/40)(40/60)(60/40). It is apparent that while there are five components of Bayes' theorem being considered (Ω_1, Ω_0, l, sample size, and conflict), these are constrained by 4 degrees of freedom. No one variable can be manipulated without changing at least one other variable. In a previous experiment (Schum & Martin, 1968), conflict was systematically varied, but by necessity the size of Ω_1, sample diagnosticity, was confounded with it. While conflict had marked effects, it was impossible to separate these effects from the strong effects of Ω_1. The present study investigated sample conflict with sample size and Ω_0 held at a constant value and with Ω_1 constrained to one of three levels.

A second purpose of the present experiment was to assess the effects of *S* experience in a decision-making context where conflict is manipulated. Some past research has exclusively employed experimentally naive *S*s (Peterson & Miller, 1965; Phillips & Edwards, 1966), while other research has used more experienced *S*s (Schum, 1966; Schum & Martin, 1968). It is thus of some importance to investigate the possible effects of experience on inferential performance.

METHOD

Subjects.—The *S*s were undergraduate students enrolled at Ohio State University. Nine naive and nine relatively experienced *S*s were used, the criterion for experience being half-time work for at least 9 mo. in multinomial probabilistic inference tasks. Naive *S*s received a 4-hr. orientation program including a general discussion of probabilities, an introduction to the task, and instructions in the use of the contingency tables and the payoff function graph.

Procedure.—A symbolic multinomial environment was used in which there were six hypotheses equally likely on an a priori basis; $\Omega_0 = 1$ for any pair. To generate a sample of data, a hypothesis was randomly chosen, a sample was drawn, and *S* was informed of the state of 6 out of 15 possible stimulus dimensions. For a given dimension, by knowing which state had occurred and referring to a contingency table available, *S* could determine which hypothesis this datum favored and by what l. For a given dimension only one l existed, but any of 6 states within this dimension was possible, each of which favored only one of the six hypotheses.

Using the urn paradigm, a dimension might be color and the state might be red. By referring to his contingency table (see Table 1), *S* might find that there were 50 red objects in a given urn and 10 in each of the other five urns ($l = 5/1$). A second dimension might be composition, with the state being metal. In this case there might be 80 metal objects in a given urn and 4 metal objects in each of the others ($l = 20/1$).

TABLE 1

EXAMPLES OF TWO POSSIBLE CONTINGENCY TABLES

Dimension and state	Urn					
	1	2	3	4	5	6
Color						
Blue	10	50	10	10	10	10
Black	10	10	10	50	10	10
Red	50	10	10	10	10	10
Green	10	10	10	10	50	10
White	10	10	50	10	10	10
Yellow	10	10	10	10	10	50
Composition						
Glass	4	4	4	4	4	80
Wood	4	4	4	4	80	4
Plastic	80	4	4	4	4	4
Metal	4	4	80	4	4	4
Paper	4	80	4	4	4	4
Plaster	4	4	4	80	4	4

In this simplified multinomial environment where a datum favors only one hypothesis over each of the others, if one ignores the size of l and considers only its direction, there are 19 ways that a sample of six data may be distributed across hypotheses. Using the definition of conflict previously given, these cases can be grouped into six conflict levels according to the number of data favoring any hypothesis other than the generating hypothesis. Five of these levels were used in the present experiment, excluding the case with no conflict.

The problems were selected such that a factorial analysis could be used with five levels of conflict and three ranges of sample diagnosticity. In terms of the size of the probability[3] under the generating hypothesis, these ranges were .45–.59, .65–.79, and .85–.99. Twenty-five samples were presented from each cell of the design, giving a total of 375 samples.

In addition to an hourly wage of $1.25/hr, *S*s received a logarithmic payoff. The payoff increased with higher probability estimates under the

---

[3] For simplicity Ω_1 has been used up to this point. However, since *S* actually gave his estimate in terms of subjective probabilities, these are used in the discussion of the payoff function and data. In the multiple hypothesis case, posterior probabilities can be converted to Ω_1 by taking the ratio of the probabilities for any two hypotheses:

$$\Omega_{1,\,A:B} = \frac{P(H_A|D)}{P(H_B|D)}.$$

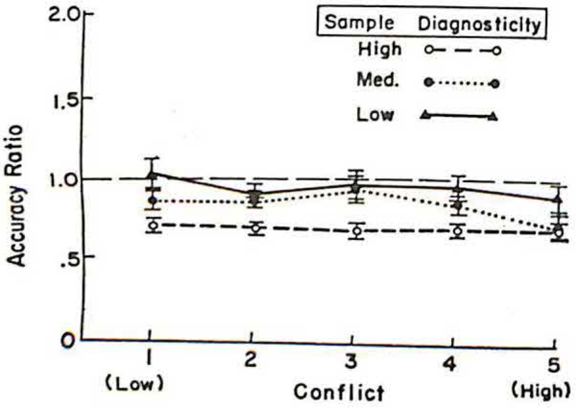


FIG. 1. Means and 95% confidence intervals for naive *S*s.

generating hypothesis, but as a logarithmically decreasing function. The function was

$$\text{payoff unit} = 10 + 12.851 \log_{10} (\text{subjective probability}). \quad [1]$$

This function has the following characteristics: (*a*) It yields no payoff if the prior probabilities are given; (*b*) at .5¢/unit, it allows a maximum of 5¢ and a minimum of −15¢ per trial; and (*c*) in order to maximize subjectively expected winnings with a logarithmic function, it has been shown that *S* should give his actual subjective probabilities rather than any other numbers (van Naerssen, 1962). Thus, a logarithmic payoff function was used in an attempt to get *S* to give his actual subjective probability estimate.

During each trial, *S* observed a closed-circuit television monitor displaying the six dimensions being reported during that trial and the particular state of each dimension. The *S* could then consult a contingency table to determine which hypothesis each datum favored and by what *l*. Within 5 min., *S* was required to record his subjective probability that each of the six hypotheses had generated this sample by writing a probability between .995 and .001 under each of the six hypotheses listed on a response sheet. It was required that this vector sum to 1.0. Payoff was calculated based on the size of the probability under the hypothesis that had generated the data using Equation 1. Payoff information was presented following each 30-trial sequence.

RESULTS AND DISCUSSION

The primary dependent variable used was one that is commonly reported in such research, the accuracy ratio, which indicates the optimality of a subjective estimate by attaching an exponent to L in the normative Bayes' model:

$$\Omega_1 = L^{AR}\, \Omega_0.$$

Thus, an $AR = 1.0$ may be considered optimal behavior; $AR < 1.0$, conservative; and $AR > 1.0$, excessive (see Schum, 1966; Schum & Martin, 1968; for a discussion of AR applied to multinomial case).

Figures 1 and 2 indicate the AR and 95% confidence intervals obtained for naive and experienced *S*s under each of the five conflict levels. In all conditions the means of the AR were ordered with high sample diagnosticity being most conservative and low diagnosticity being least conservative. Across conflict levels and *S*s the means of the AR for sample diagnosticity levels were: .626 (high diagnosticity), .803 (medium), and .979 (low).

The analysis of variance for AR showed that the main effects of diagnosticity level, $F\ (2,\ 32) = 71.79$, $p < .001$, and conflict level, $F\ (4,\ 64) = 17.27$, $p < .001$, were significant. The main effect of experience was not significant. All interactions were significant at $p < .001$ except Diagnosticity Level × Experience, which was significant at $p < .01$.

A second dependent variable was also examined: the number of times *S* assigned his highest posterior probability estimate to the hypothesis that actually generated the sample. Since *S* did not actually choose a single hypothesis as correct, this measure is merely an estimate of how often he would have been correct if he had been required to choose one and had behaved in accordance with his probability vector. Thus, this measure indicates how large an effect conflict

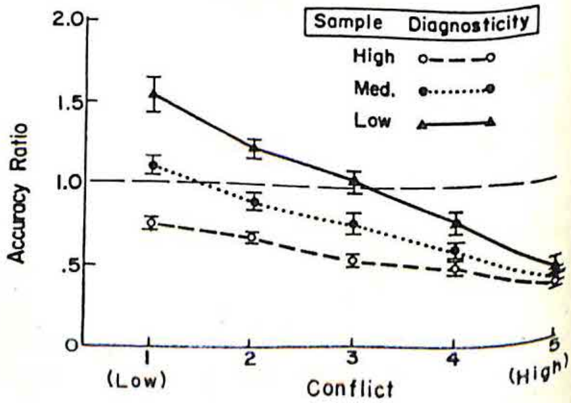


FIG. 2. Means and 95% confidence intervals for experienced *S*s.

might have in a situation where *S* simply chooses the hypothesis he believes generated the data sample.

The mean number of correct endorsements in each condition is presented in Fig. 3 and 4. Sample diagnosticity had a marked effect on this variable. With two exceptions, the means are ordered with high diagnosticity associated with more correct endorsements than low. The analysis of variance for correct endorsements indicated that the main effects of diagnosticity level, $F\ (2, 32) = 34.9$, $p < .001$, and conflict level, $F\ (4, 64) = 12.9$, $p < .001$, were significant, while that of experience was not, $F\ (1, 16) = .58$, $p > .25$. All interactions except Conflict Level × Diagnosticity Level were also significant.

The most striking result is the overall superiority of naive *S*s over experienced *S*s. While the expected effect of higher sample diagnosticity being associated with lower *AR* was found for both groups of *S*s, the effect was significantly less pronounced for naive *S*s, as shown by the significant Diagnosticity Level × Experience interaction, $F\ (2, 32) = 5.54$, $p < .01$. This result would indicate that the naive *S*s were more appropriately weighting the total impact of the sample of data than the experienced *S*s.

Naive *S*s were also superior in handling the range of conflict used. Whereas the *AR* of naive *S*s remained stable and close to optimal across conflict levels, experienced *S*s became generally less optimal with higher conflict. The Conflict Level × Experience interaction indicates that this effect was significant, $F\ (4, 64) = 11.43$, $p < .001$. In terms of the mean number of correct implied endorsements, conflict level again had a marked effect on experienced *S*s' behavior and little effect upon naive *S*s' behavior. At lower levels of conflict, experienced *S*s seemed to perform somewhat better than naive *S*s, but as conflict level increased, experienced *S*s rapidly lost this advantage. Again, this effect was significant as indicated by the Conflict Level × Experience interaction, $F\ (4, 64) = 8.47$, $p < .001$.

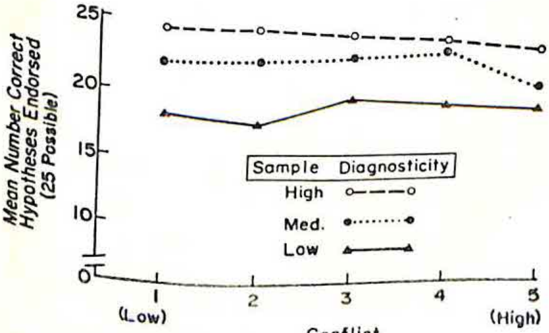


FIG. 3. Mean number of correct endorsements for naive *S*s.

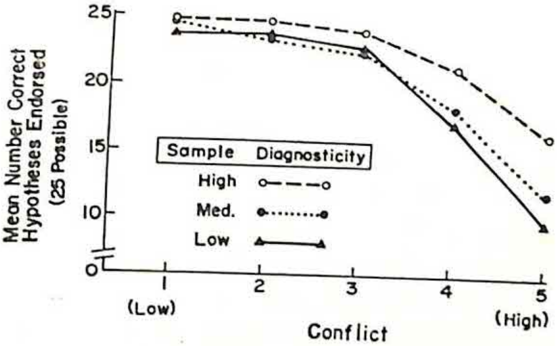


FIG. 4. Mean number of correct endorsements for experienced *S*s.

The prediction that the overall performance of experienced *S*s would be superior to that of naive *S*s due to a positive general transfer of training from previous inferential tasks was not supported. A possible explanation for the superior performance of naive *S*s is that a negative specific transfer occurred for experienced *S*s. Rather than a transfer of an appropriate general inference behavior such as an intuitive knowledge of Bayes' theorem, what may have occurred is the inappropriate transfer of specific response tendencies. If this explanation is valid, much of the research which has been carried out using experienced *S*s must be evaluated in terms of the specific characteristics of the task used to train *S*s. Clearly, it is of importance to specify the procedure for training decision makers if one wishes to generalize experimental results.

The direct relationship found between conservatism and conflict for experienced *S*s could be attributed to an inference behavior in which *S* gives some constant weight to a datum regardless of its *l*. He might perceive the size of the *L* inappropriately by attaching too great an importance to the number of data favoring each hypothesis and too little importance to the weighting of each datum.

The present results indicate several important considerations for the use of man as a decision maker. It would seem to be of some importance to know generally how diagnostic the evidence toward a certain hypothesis is likely to be in the decision-making situations that he will be called on to face. Man seems to be only slightly conservative when a sample of data indicates the truth of a hypothesis to only a moderate degree. When data are highly diagnostic, man appears considerably less optimal. Second, it is important to know the amount of conflict likely to be encountered in a sample of evidence. The present results indicate that man may place too much emphasis on the number of supporting data and too little upon the appropriate impact of each datum. Third, type of training for a decision maker seems to affect the kind of decisions he will make. Whenever possible, the training task for decision makers should incorporate parameters constrained to ranges likely to be encountered in actual decisions.

REFERENCES

PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, **68**, 29–46.

PETERSON, C. R., & MILLER, A. J. Sensitivity of subjective probability revision. *Journal of Experimental Psychology*, 1965, **70**, 117–121.

PETERSON, C. R., SCHNEIDER, R. J., & MILLER, A. J. Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology*, 1965, **69**, 522–527.

PHILLIPS, L. D., & EDWARDS, W. Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 1966, **72**, 346–354.

SCHUM, D. A. Prior uncertainty and amount of diagnostic evidence as variables in a probabilistic inference task. *Organizational Behavior and Human Performance*, 1966, **1**, 31–54.

SCHUM, D. A., & MARTIN, D. W. Human processing of inconclusive evidence from multinomial probability distributions. *Organizational Behavior and Human Performance*, 1968, **3**, 353–365.

VAN NAERSSEN, R. F. A scale for the measurement of subjective probability. *Acta Psychologica*, 1962, **20**, 159–166.

(Received October 24, 1968)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 9–13

PERCEPTION OF THE TWO-PRONGED TRIDENT BY TWO- AND THREE-DIMENSIONAL PERCEIVERS

J. B. DEREGOWSKI[1]

Institute for Social Research, University of Zambia, Lusaka

The *S*s were categorized as either two-dimensional (2D) or three-dimensional (3D) perceivers depending on whether they built a 2D or a 3D model in response to a drawing. An hypothesis that there would be no difference between the two groups when asked to copy an "impossible" figure under conditions requiring little short-term memory storage, but that the 3D group would find the task more difficult if such storage was required was investigated and confirmed. No confirmation was obtained for another hypothesis that the absence of background would make the task of copying the impossible figure more difficult. The results suggest that the quality of a reproduction cannot be used as an indicator of presence or absence of 3D pictorial perception and confirm that the difficulties in interpretation of pictorial depth cues are experienced by some *S*s.

The *S*s drawn from certain cultural groups that have difficulties with three-dimensional perception of pictorial material should not find the "impossible" figures (Gregory, 1966) confusing, for the confusion only arises if one attempts to interpret a pattern of lines as depicting a three-dimensional object. When such an attempt is not made, any pattern of lines should be accepted as satisfactory. Deregowski (1968b) used the task of copying a two-pronged trident (the procedure used being similar to that described subsequently) in his experiment into depth perception in Zambian children, and the results obtained supported the aforementioned hypothesis. However, his subsequent replications suggested that *S*s performing the task may find the two-pronged trident confusing only if a delay occurs between the examination of the drawing and the beginning of reproduction by *S*, a variable that was not considered in the original experiment. Such a suggestion gains support from Attneave's (1955) experiment, wherein it was found that for the same information content, a delayed reproduction caused much greater divergence between responses to random and symmetrical stimuli than did immediate reproduction. That is, stimuli that were difficult to organize perceptually were also found to be harder to reproduce; and this difficulty increased with the time interval separating presentation from reproduction.

It has been suggested (Maistriaux, no date; Deregowski, 1968b) that the presence of background may be a contributory factor to the difficulties of pictorial perception. Assuming that the presence of a background, which does not incorporate additional depth cues, tends to make two-dimensional perception more probable, one would expect its removal to lead to three-dimensional perception and hence to greater confusion when the stimuli represent "impossible" figures.

The following hypotheses were therefore postulated:

Hypothesis 1. There will be a significant difference between two-dimensional (2D) and three-dimensional (3D) *S*s when copying a two-pronged trident only when a period of delay between inspection of the stimulus and drawing response is introduced. The 3D *S*s will be found to have greater difficulties.

Hypothesis 2. The difficulties of *S*s responding to the confusing figure will be greater when no stimulus background is visible than when such background is visible.

METHOD

Experimental Plan

Each *S* was classified as a 2D or a 3D perceiver on the basis of his response to the construction tasks. He was subsequently allocated to one of the four possible combinations of the following experimental conditions: duration of delay (.1 or 10.0 sec.) with either the luminous models or the drawings being used as stimuli.

---

[1] Requests for reprints should be sent to J. B. Deregowski, who is now at the University of Aberdeen, Aberdeen, Scotland.

Construction Tasks

Materials.—The following materials were provided for all tasks involving construction: 14 plasticine balls, ½-in. diameter; bamboo splints varying in length by 1-in. steps from 2½ in. to 7½ in. and forming six sets of 12 splints each, the splints within each set being equally long.

A preliminary construction task was used to introduce *S* to the construction materials provided. A tetrahedron consisting of four plasticine balls connected by six 7-in. bamboo splints was placed on a table in front of *S*. The model was pointed out to *S*, and he was requested to build a "thing" like it using wood and "clay" (plasticine), which were also pointed out to him. When *S* had indicated that he had completed his model, it was taken to pieces and *S* was given his first test task.

Construction Test

Materials.—The construction materials previously described were used with a set of drawings (see Fig. 1) that were intended to embody the following depth cues: (i) overlap, (ii) size difference, and (iii) size difference and overlap.

Procedure.—The drawings were presented vertically in the aforementioned sequence in front of *S*, who was instructed: "I want you to build what you see on the picture." Any further inquiries were met with: "Just as you see on the picture." When *S* indicated that he had completed his model, *E* judged this model to be 2D or 3D, recorded his judgments, and made a sketch of the model. The model was then taken to pieces and the next drawing was presented with the same instructions.

The criteria for 3D and 2D judgments were as follows: A response was judged to be 3D if all the members did not lie in parallel or nearly parallel planes separated by a distance not exceeding ½ in.; all other responses were classified as being 2D. No difficulty was experienced in applying this criterion as no models fell within the borderline region.

Any *S* who made a 3D response to all three drawings was classified as a 3D *S*; all other *S*s were classified as 2D.

Drawing Task

Materials.—The apparatus was placed in two neighboring cubicles having an interconnecting window. One of the cubicles was used to display the stimuli while *S*s sat in the other. The *S*s could see the displayed stimulus by lifting a 13 × 8 in. flap in the screen obscuring the window. Lifting of the flap started a stop clock (Spencer, 1965) that measured the duration for which the window was uncovered in .1-sec. units. Closure of the flap started a Birkbeck timer, which, depending on the setting used, produced an auditory signal either .1 sec. or 10.0 sec. after closure. The display cubicle contained either three flat wire models painted with paint which reradiated ultraviolet light in visible spectrum and an ultraviolet light source or a set of three drawings (see Fig. 2) illuminated by an ordinary incandescent lamp. The models glowed orange-yellow when exposed to ultraviolet light in the otherwise dark cubicle. The drawings, which were of the same size and shape as the models, were drawn on white cardboard in India ink. Both drawings and the models were suspended in a triangular formation, the introductory stimulus (see Fig. 2a) being above the other two stimuli. The two-pronged trident hung to the left of the trident. A set of cardboard screens enabled any desired stimulus to be exposed at any time.

Procedure.—Each *S* was shown first how the flap could be lifted. Then he was given a sheet of foolscap paper and a ball-point pen and asked to draw what he saw through the window, the introductory model being on display. He was told that he could lift the flap as often as he wished and look at the stimulus for as long as he wished, but that he must not draw while the flap was open and must wait for the signal (this was demonstrated) before beginning to draw. During this task, any misunderstandings (such as the tendency of some *S*s who responded to drawings to draw the screens concealing the other stimuli) were resolved. Subsequently, the introductory stimulus was covered, one of the other stimuli was displayed, and *S*

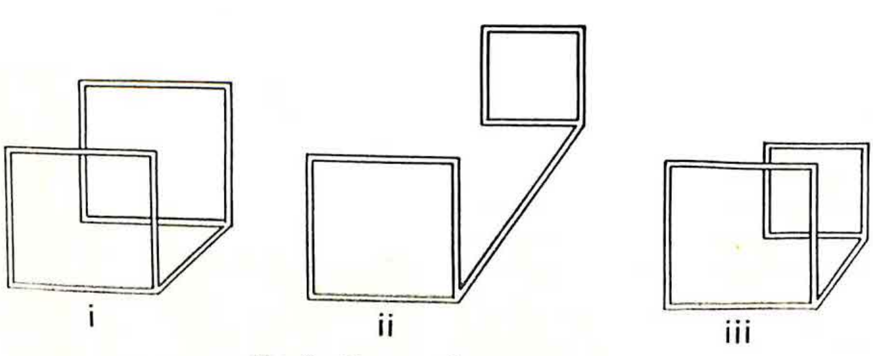


FIG. 1. Construction test stimuli.

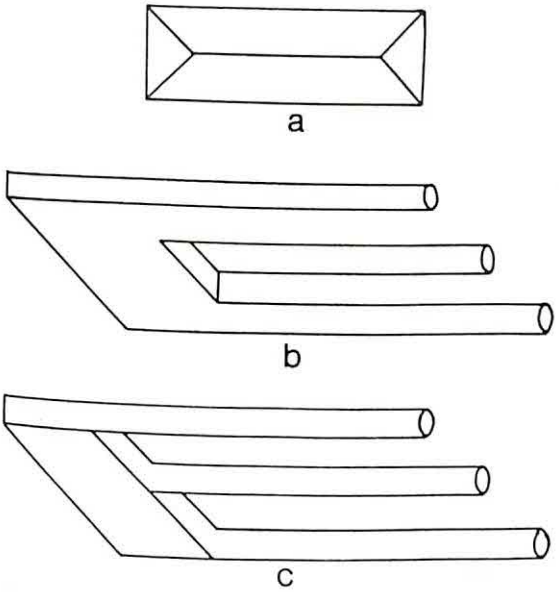


FIG. 2. Drawing task stimuli.

was asked to draw it. The order of the test stimuli was alternated so that half of the *S*s within each group responded to each of the two possible sequences.

As soon as *S* indicated that he considered his drawing to be finished, the total time he kept the flap open was recorded.

Subjects.—Sixty-seven boys were drawn from the third and fourth grades of two non-fee-paying primary schools in Lusaka. The sample characteristics, which excluded 3 rejected *S*s, were: 30 Grade 4 *S*s, 34 Grade 3 *S*s; mean age = 10.7 yr, *SD* = 1.9. The age parameters are based on 63 *S*s, as 1 *S* did not know his age.

RESULTS

Three *S*s had to be rejected, two because, contrary to the instructions, they included irrelevant material in the test reproductions and the third because he constructed tetrahedra (a type of response appropriate to the introductory construction task) in response to all three of the construction test drawings. The mean durations of inspection times of the acceptable *S*s are given in Table 1. These were not used directly in the analysis of data; instead, for each *S* a ratio of the difference between the time spent looking at the two-pronged trident and the time spent looking at the trident to the sum of these times was calculated. This was termed coefficient *C* and was done to reduce inter-*S* differences, which might have roots in the differences in ability to draw and in the standard of perfection aimed at. The higher the value of this coefficient, the more difficult does *S* find the copying of the confusing figure.

These ratios (for mean values, see Table 1) were used in an analysis of variance. The only main effect approaching significance is the delay between the closing of the flap and the signal to draw, $F\ (1, 56) = 3.20$, $p < .10$. The second-level interaction is nonsignificant, but one of the first-level interactions, that between the category of the *S*s and the duration of delay, is $F\ (1, 56) = 5.75$, $p < .025$. Tukey's tests, therefore, were applied, and it was found that for both background conditions, there exist significant differences in the predicted directions between the responses of 3D *S*s in .1-sec. and 10.0-sec. delay conditions ($p < .01$ for nonbackground conditions; $p < .05$ for background conditions), as well as between the responses of 2D *S*s and 3D *S*s in 10.0-sec. delay conditions ($p < .05$). These relationships find confirmation in the comparison of the frequencies with which the responders in the combined background and nonbackground groups fell into two categories, those supporting the postulated hypothesis and those failing to support it. Application of Fisher's exact test to the data (see Table 1) yields significant results both in the case of the comparison of the responses of 3D *S*s with those of 2D *S*s in 10.0-sec. delay conditions ($p < .01$) and in the case of the comparison of the 3D *S*s' responses in .1-sec. and 10.0-sec. conditions ($p < .025$).

There is also a single statistically significant difference (Tukey's test, $p < .05$) in the .1-sec. delay condition. This is the difference between the two nonbackground groups, the 2D group having a higher coefficient *C* value than the 3D group.

Three independent judges allocated drawings obtained in response to each of the experimental stimuli into three categories: good, indifferent, and bad. The data thus gained were used for the following comparisons: (*a*) Within each judge/stimulus combination, the frequency of the response of the 3D *S*s was compared with that of the 2D *S*s. In no case did the difference approach statistical significance. (*b*) Those "bad" copies

TABLE 1

SUMMARY OF OBTAINED DATA

Type of *S*s	Data	.1-sec. delay		10.0-sec. delay	
		No background	Background	No background	Background
2D	Raw scores (in sec.)				
	Mean values for TPT	18.5	21.4	28.6	41.2
	Mean values for T	14.5	27.1	30.4	44.2
	Coefficient *C* (*M*)	.703	−.350	−.015	−.175
	No. of *S*s taking longer to copy TPT	5	3	3	2
	No. of other *S*s	3	5	5	6
3D	Raw scores (in sec.)				
	Mean values for TPT	13.7	17.7	22.1	26.0
	Mean values for T	16.2	19.4	17.0	21.0
	Coefficient *C* (*M*)	−.650	−.333	1.568	1.183
	No. of *S*s taking longer to copy TPT	3	3	6	7
	No. of other *S*s	5	5	2	1

Note.—TPT = two-pronged trident figure; T = trident figure.

of the confusing drawing on which all three judges agreed were analyzed to see whether the difference between the number of 2D and 3D responders within the category differed from chance. It was found that the category contained 15 2D and 15 3D *S*s. The other categories were too small to permit such analysis.

DISCUSSION

The results support the first hypothesis. It seems that when very little retention is required, *S*s from both populations experience no special difficulties with the two-pronged trident. However, when *S*s are required to remember the stimuli, even for a relatively short time, the 3D *S*s, who presumably find the two-pronged trident confusing, find its reproduction significantly more difficult than do the 2D *S*s. This suggests that depth perception once acquired cannot be easily suppressed, for if it could, the 3D *S*s would have probably taken advantage of this. As a further support for the previous suggestion can be adduced the absence of 2D/3D borderline responses in the construction task, which was observed both here and previously (Deregowski, 1968a, 1968b), as well as the consistency of *S*s in making one of the two possible types of response. Thus of the 64 *S*s tested, only 6 gave responses that were not either all 3D or all 2D.

It must be noted that the 2D/3D dichotomization used here is not synonymous with the division of *S*s that can be arrived at on the basis of Hudson's (1960, 1967) test, for it has been shown (Deregowski, 1968a, 1968b) that the construction test differs from Hudson's test by evoking a larger number of 3D responses. In this sense, therefore, pictorial depth perception is not a yes/no phenomenon; Hudson's test pictures, which are meaningful representations of a hunting scene, are, however, of an entirely different nature from the stimuli used here. It is difficult, therefore, to say to what extent the two tests sample the same phenomena.

Insofar as the construction test and the drawing task employed here reinforce each other, the results suggest that a proportion of the population sampled is incapable of perceiving depth in simple geometrical figures.

The second hypothesis did not gain support from the data. In the case of the stimuli employed, the background appears to play no significant part in perception of the stimuli used. It may be, however, that it is of consequence in a different type of pictures.

It appears that the drawings of the 3D *S*s do not differ significantly from those obtained from the 2D *S*s and hence that the quality of reproduction is not affected by the perceptual style of *S*. Given sufficient time, both groups are capable of producing drawings of comparable quality.

Examination of the drawings reveals a large range of accuracy of representation extending from correct copy to a highly impressionistic depiction of the trident representing an object with as many as six prongs. These fantastic distortions probably correspond to similar distortions in response to the construction task observed both here and previously (Deregowski, 1968a, 1968b). Further, the drawings are frequently both oddly oriented (the trident is often rotated through 90°) and oddly placed relative to the sheet on which they were drawn. This is probably the same phenomenon described by Hudson (1962, 1967) and not specially related to the task used here.

REFERENCES

ATTNEAVE, F. Symmetry, information and memory for patterns. *American Journal of Psychology*, 1955, **68**, 209–222.

DEREGOWSKI, J. B. Difficulties in pictorial depth perception in Africa. *British Journal of Psychology*, 1968, **59**, 195–204. (a)

DEREGOWSKI, J. B. Investigation into pictorial perception with special reference to Central Africa. Unpublished doctoral dissertation, University of London, 1968. (b)

GREGORY, R. L. Visual illusions. In B. M. Foss (Ed.), *New horizons in psychology*. Harmondsworth, Middlesex: Penguin Books, 1966.

HUDSON, W. Pictorial depth perception in subcultural groups in Africa. *Journal of Social Psychology*, 1960, **52**, 183–208.

HUDSON, W. Pictorial perception and educational adaptation in Africa. *Psychologia Africana*, 1962, **9**, 226–239.

HUDSON, W. The study of the problem of pictorial perception among unacculturated groups. *International Journal of Psychology*, 1967, **2**, 89–107.

MAISTRIAUX, R. *L'intelligence noire et son destin*. Bruxelles: Les editions de problemes d'Afrique Centrale, no date.

SPENCER, J. An inexpensive digital stop-clock. *British Journal of Psychology*, 1965, **56**, 317–319.

(Received December 13, 1968)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 14–21

RECALL FOR ORDER AND CONTENT OF SERIAL WORD LISTS IN SHORT-TERM MEMORY[1]

ALFRED H. FUCHS[2]

Bowdoin College

The role of serial order in recall was studied with the use of an order cue to test for recall of words at each of five serial positions. Word strings received one, two, or three presentations, and recall was tested after 4, 8, or 16 sec. Analyses of errors revealed that order errors could be separated from errors of content primarily because the former showed a serial position effect, while the latter did not. Order is an aspect of a stimulus series that may serve as a means of organizing content in memory.

Ordered recall of serially presented items may result in the correct recall of items in a partly or wholly incorrect sequence. The problem of assessing the independent effects of item and order information in recall is difficult (Brown, 1958; Crossman, 1960); "if any item information is lost then the assessment of order information cannot be made at all because the only data on which it can be based will have gone [Conrad, 1959, p. 352]." The conditions under which the loss of order information can be assessed are those that result in the correct recall of items in a transposed sequence. Because transposition errors in letter series were more probable when acoustically similar letters were in adjacent positions, Conrad (1965) argued that it was both unnecessary and untenable to assume the independence of content and order information in determining such errors. The basis of transposition errors was thus assumed to be a failure to discriminate among similar items during recall, with acoustically similar letters in adjacent positions transposing more readily and providing spurious evidence of order error.

However, long-term memory (LTM) research contains evidence of intrusions from previous lists (e.g., Melton & von Lackum, 1941). Similar errors have been observed in repeated tests of immediate memory (Conrad, 1959) and in short-term memory (STM) (Fuchs & Melton, 1964). In these experiments, elements that appeared in a particular serial position as a stimulus tended to occur as intrusions in a subsequent list, if they occurred at all, in the same serial position. This evidence suggests that order is an aspect of an element in memory that can serve as a dimension of similarity among elements and influence recall error. Therefore, it may be premature to assume that order error is simply an artifact of item error.

The present study was designed to investigate serial order through the use of a probe test for recall of single words from a string of five words. Words were chosen because their uniqueness would permit the tracing of errors across word groups, and five-word strings were used to approach the limit of the memory span. Frequency of presentation was chosen as a within-group variable since previous research (Fuchs & Melton, 1964) had shown that serial position intrusions were more likely to appear in strings presented only once but to derive from prior strings which had been presented more frequently.

METHOD

Apparatus.—A 1-ft. square, ground-glass screen was mounted in a plywood sheet (about 3×4 ft.) fastened at right angles to a table top. Slides were projected from behind; *S* sat approximately

---

[1] An earlier version of this paper was presented at the meeting of the New Zealand Psychological Society, Auckland, August 1968. The research was supported in part by Grant GE. 2903 from the National Science Foundation. This report was prepared while the author served as visiting professor at the Victoria University of Wellington, New Zealand; the author is grateful to the Department of Psychology for aid in the preparation of the figures and is indebted to L. B. Brown for critical and helpful discussions of the manuscript.

[2] Requests for reprints should be sent to Alfred H. Fuchs, Department of Psychology, Bowdoin College, Brunswick, Maine 04011.

4 ft. from the screen. Hidden from *S*'s view were the Kodak Carousel slide projector, a Gerbrands voice key and Hunter Klockounter for obtaining recall latencies,[3] and a Gerbrands VI timer for programming the slides and starting the Klockounter. Tape-recorded white noise was played through earphones to *S* to mask apparatus noise during the experiment.

Subjects.—The *S*s were 49 male, paid, senior students. Three were dropped because of apparatus failures during their experimental session, and one was eliminated for failing to follow instructions.

Stimuli.—The stimuli employed were 250 four-letter English nouns. No words beginning with Q, U, or X were used, and all words were selected from the Thorndike-Lorge (1944) word count within the range of 1–100 occurrences per million. The words selected were monosyllabic and as nonhomophonic as possible.

The stimulus words were arranged into strings of five words, with the following restrictions: (*a*) No two words of the group had the same first letter; (*b*) no word began with the letter with which the previous word ended; (*c*) obviously meaningful combinations of words were eliminated, e.g., BARN DOOR or BITE MEAT. The words were originally selected and employed for a similar purpose by Noyd (1964) and were those used by Fuchs and Melton (1964).

Procedure.—Instructions were tape-recorded and played to *S* via earphones; *S* was told that the experiment was concerned with memory for words and "information processing" (counting backward). It was hoped that this emphasis on the task of backward counting would encourage *S* to treat this task seriously to minimize rehearsal during the retention interval. The *S* was instructed to read the stimulus words aloud as often as they appeared, to count backward aloud by threes from the three-digit number when it appeared, and to recall as quickly and accurately as possible when the relevant cue appeared.

The first slide contained five words arranged in a "stair-step" fashion from the upper left corner to the lower right corner of the slide. Each word was positioned in the space between 3-in. horizontal lines drawn on the ground-glass projection screen. Stimulus slides appeared for 4 sec., were presented one, two, or three times, and were followed by a slide containing a three-digit number to initiate backward counting for the duration of the retention interval.

The recall cue was an asterisk in one of the spaces in which the words had been positioned on the screen. The task for *S* was to recall the word that had occupied that space. The duration of the recall slide was 8 sec., which was followed by a 4-sec. "rest" slide containing a matrix pattern.

Experimental design.—There were three experimental groups, differentiated by the length of the retention interval for each (4, 8, or 16 sec.). Assignment of *S*s to groups was done according to a prearranged sequence, with each group receiving 5 *S*s, randomly assigned, of each 15 who appeared until each group was complete. Within-group variables included frequency of presentation and the serial position tested on any trial.

Within each block of 15 trials, there was 1 trial for each of the three presentation frequencies combined with each of the five serial positions. The sequence of trials was randomized within trial blocks of 15; each group of *S*s received the same sequence of trials. A total of 50 trials was presented; the first 5, unknown to *S*, were practice trials in which each serial position was tested once. Accuracy and speed of recall were recorded by *E* on each trial.

RESULTS

Correct recall.—Each of the five serial positions was tested three times under each presentation condition; the basic datum per *S* was therefore a score from 0 to 3. An analysis of variance of the correct recall data showed that effects of retention interval (groups), $F\ (2, 42) = 7.04$, presentation frequency, $F\ (2, 84) = 122.71$, and serial position, $F\ (4, 168) = 36.30$, were statistically significant ($p < .01$). The only significant interaction among these variables ($p < .01$) was the Presentation Frequency × Serial Position interaction, $F\ (8, 336) = 10.60$; the others were not significant ($p > .10$). Summarized across retention intervals of 4, 8, and 16 sec., total correct were 430, 336, and 327, respectively, of a total possible correct at each interval of 675 (15 *S*s × 45 trials).

The distribution of correct responses across serial positions for each of the presentation frequencies for each retention interval is shown in Fig. 1. The largest decrease in correct recall is between 4 and 8 sec.; at each retention interval there are large differences attributable to the effects of repetitions. Serial Position 5 shows dramatic improvements at each retention interval when presentations are increased from 1 to 2, with relatively little change from 2 to 3. Serial Position 1, on the other hand, remains relatively constant over increased presentation frequencies. The possibility that the constancy of recall at Serial Position 1 was the source of the significant interaction was tested by a second analysis of variance with Serial Position 1 data omitted; the interaction was still significant, $F\ (6, 252) = 5.14$, $p < .01$, and indicates that the rates of improvement in recall at each of the serial positions differ as presentations are increased. That presentation frequency and serial position do not interact with retention interval suggests that the effects of the latter are to reduce the magnitudes of the frequencies correct at each serial position and presentation frequency with relative uniformity as the retention interval is lengthened.

---

[3] Latency data may be obtained in a mimeographed report from the author; these data supplement the error data and their interpretation, but in the interests of space are not reported here.

Errors.—The incorrect responses were sorted into four error types: (*a*) order errors, which consisted of words intruded from a serial position within a five-word string in place of the word from the position being tested; (*b*) intrusions from a previous list, which were words from a preceding five-word string intruded as errors in recall of a subsequent string; (*c*) extralist intrusions, which constituted intruded words not employed in the experiment; and (*d*) omissions, which included both "don't know" and failures to respond within the 8-sec. recall period. Order errors were most frequent and accounted for 521, or 55.9% of all errors; omissions numbered 184, or 19.7%; intrusions from previous lists were 152, or 16.3%; and extralist errors numbered 75, or 8.0%. The last two categories of errors are comparable to Conrad's (1959) serial order error and substitution error.

Since an analysis of total error frequencies would have duplicated that for the correct responses, separate analyses were carried out for (*a*) order errors and (*b*) all others. The analysis of variance for order errors revealed the same pattern as that for the correct response data: presentation frequency, $F\ (2, 84) = 5.82$, serial position, $F\ (4, 168) = 26.30$, and their interaction, $F\ (8, 336) = 2.71$, were all significant ($p < .01$); but retention interval was not statistically significant, $F\ (2, 42) = 1.16$, $p > .10$.

The immediate consequence of this failure to reject the null hypothesis was to permit the data to be summarized across retention interval. These data are plotted in Fig. 2, which shows the percentage of responses from each serial position as a function of the serial position under test for each presentation frequency. The percentages were calculated by constructing a matrix of the serial positions of responses (rows) and the serial positions under test (columns), summing the total responses for columns (serial posi-

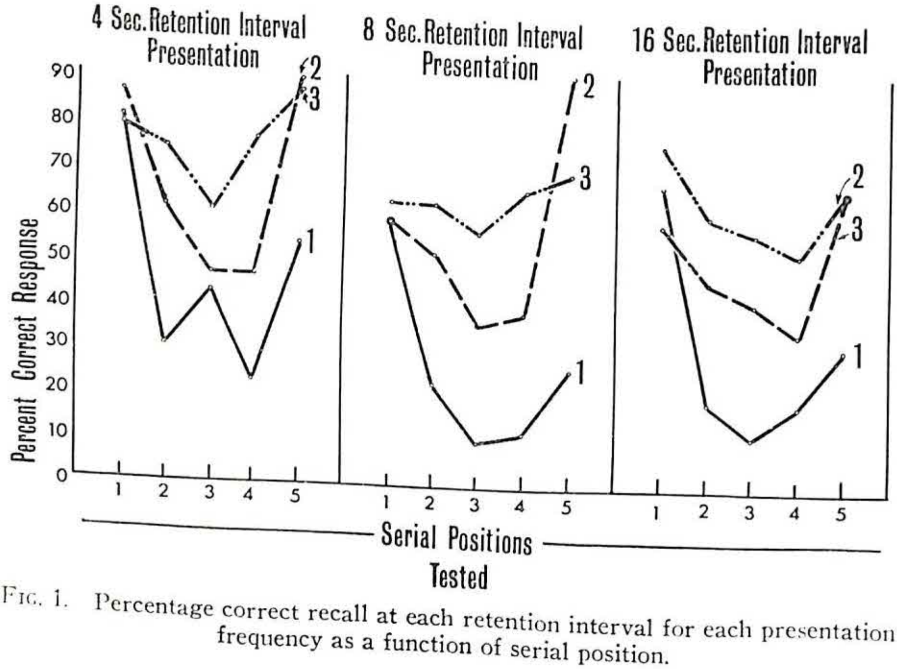


FIG. 1. Percentage correct recall at each retention interval for each presentation frequency as a function of serial position.

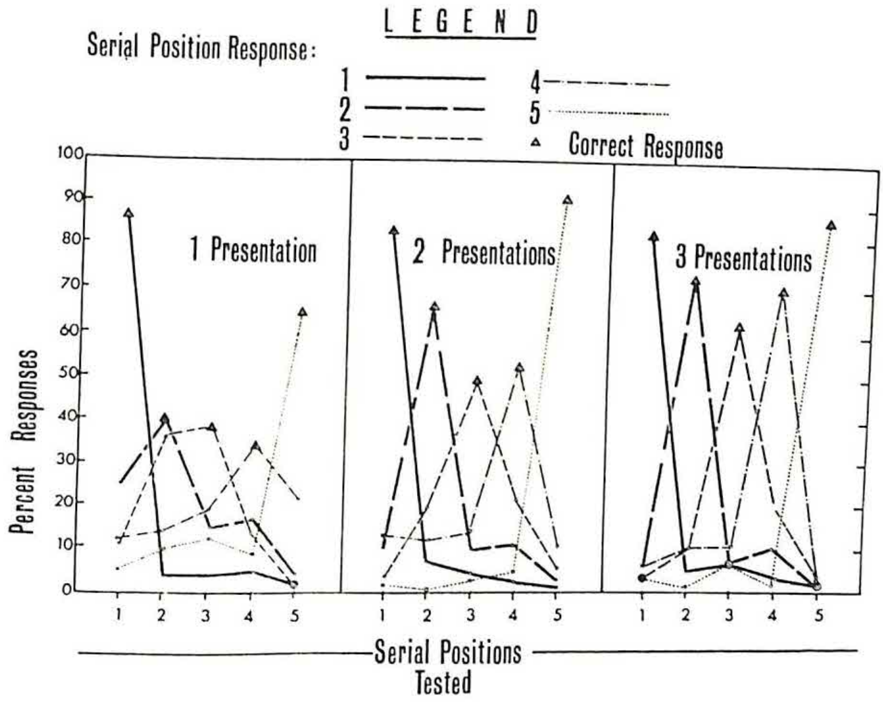


FIG. 2. Percentage of responses emitted from each serial position as a function of the serial position tested for each presentation frequency.

tion tested), and dividing each cell frequency in that column by the column total to obtain the percentage of responses from each serial position emitted when a given serial position was tested. Thus, the solid curve representing Serial Position 1 with a presentation frequency of 1 reveals that of all responses emitted when Serial Position 1 was tested, 86.4% were words correctly recalled from Serial Position 1, 3.6% were words from Serial Position 2, 3.6% were words from Serial Position 3, 4.5% from Serial Position 4, and 1.8% from Serial Position 5. Clearly, when Serial Position 1 was tested, the probability of getting the correct word was high, with the errors about equally distributed across the remaining serial positions. The curve for Serial Position 1 responses remains relatively constant across the presentation frequencies. The curve for Serial Position 5 is analogous except for the large increase in correct responding as presentations were increased from 1 to 2.

The curves for the remaining serial positions become increasingly leptokurtic with increased presentations because correct responses become more frequent and errors more distal to the position tested become less frequent than responses proximal to the position under test. This suggests that recall improves as a function of increased stimulus presentations because of improved learning of the sequence of words. Location information of words in the middle is learned more slowly, but for all positions, errors from distal positions are more rapidly eliminated than errors from adjacent positions.

The remaining errors (omissions, extralist, and intrusions from previous lists) were pooled and subjected to an analysis of variance; unlike the result for order errors, retention interval was again significant, F (2, 42) = 9.11, $p < .01$, and like both previous analyses, presentation frequency, F (2, 84) = 7.25, serial position, F (4, 168) = 24.9, and their interaction, F (8, 336) = 4.78, were statistically significant ($p < .01$).

The legitimacy of grouping all these errors together and treating them as a homogeneous group might be questioned; to treat each separately would have been difficult because of the number of empty cells due to correct responses and order errors.

As the results suggest, however, the similarities among these errors would seem, for present purposes, to outweigh possible differences.

The first characteristic of these errors is their increase with increased retention interval; Fig. 3 shows that these errors are relatively homogeneous with respect to this variable. The summary curve can be contrasted with the curve for order errors; both increase from 4 to 8 sec., but order errors show a decrease from 8 to 16 sec., while nonorder errors continue to increase. This shift in error type as a function of retention interval length may provide indirect evidence for a change in response criteria (Murdock, 1966).

The relations of presentation frequency and serial position to these errors are somewhat different than for order error. Of particular interest are the intrusions from previous lists; prior results suggest that intrusions of a previously presented word will occur in the same serial position the word had occupied when it was a member of a stimulus string. The relationship of the serial position of the intruded word (error) when it was a member of a stimulus string and the position in which it appeared as an error is shown in Table 1.

The error frequencies show no serial position effect, except that Position 4 contains the largest number of errors. The distribution of the errors in Table 1 roughly conforms to the expectation that the initial position of a word is related to where it will be subsequently intruded. Serial Positions 1, 4, and 5 show this effect most clearly; the distribution of responses across Position 3 shows a maximum concentration at Positions 3 and 4, while intrusions at Position 2 are more evenly derived from Positions 2, 3, and 4.

TABLE 1

FREQUENCY OF SERIAL POSITION ERRORS IN RELATION TO THE SERIAL POSITION OF A WORD IN A SOURCE STRING AND ITS OCCURRENCE AS AN ERROR IN A SUBSEQUENT TEST OF A DIFFERENT STRING

Serial position in source	Serial position tested				
	1	2	3	4	5
1	18	4	4	8	2
2	5	7	3	5	7
3	3	8	8	4	5
4	1	5	12	16	3
5	1	1	2	6	14

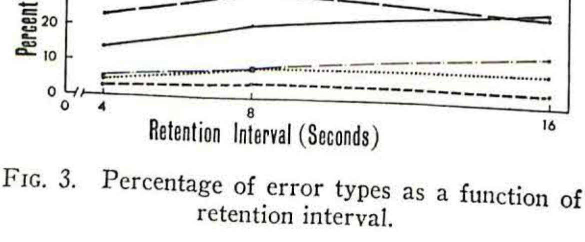


FIG. 3. Percentage of error types as a function of retention interval.

The number of repetitions given the stimulus from which an intrusion derived was obtained for each intrusion from a previous list; 31, 53, and 68 intrusions came from stimulus strings presented one, two, or three times, respectively. Intrusion frequency did not vary systematically with serial position. The uniqueness of words as stimuli in this regard, compared to the limited vocabulary size of letters and numerals, made such a tabulation possible. When stimulus lists of unequal presentation are to be recalled, intrusions from lists of greater presentations are probable in lists of fewer presentations. The position of an intrusion will depend on its position in the list from which it comes. Such errors may derive from LTM (Craik, 1968) or may represent evidence for proactive inhibition (Keppel & Underwood, 1962).

The occurrence of intrusions from serial lists combined with omissions and extralist intrusions across serial positions for each frequency of repetition is shown in Fig. 4. The clear serial position effect, apparent for order errors in Fig. 2, is absent here; furthermore, the relatively dramatic decrease in errors with repetitions contributes heavily to the increase in correct responding noted in Fig. 1 and 2. The lack of the serial position effect in Fig. 4 indicates that these errors are relatively independent of serial position and that the serial position effect is largely attributable to order errors, as shown in Fig. 2, although there is a strong primacy effect evident in Fig. 4 at 1 presentation.

"Good" versus "poor" subjects.—Conrad (1959) reported that *S*s who make relatively few errors have a higher proportion of transposition errors than *S*s who make large numbers of errors. The observation implies that "good" *S*s recall the mnemonic content but show *order* error, while "poor" *S*s make *content* errors. In the present experiment the proportion of correct responses, order errors, and other errors were determined for the five *S*s with the most errors and the five *S*s with the least errors in each group. The results are shown in Table 2.

At all retention intervals and for all frequencies of presentation, good *S*s recall more words correctly than poor *S*s, but the differences become larger as retention interval increases. In all cases, as frequency of presentation increases, the proportion of order error decreases, except for poor *S*s at 16 sec. for whom increased presentations resulted in increased order error responses.

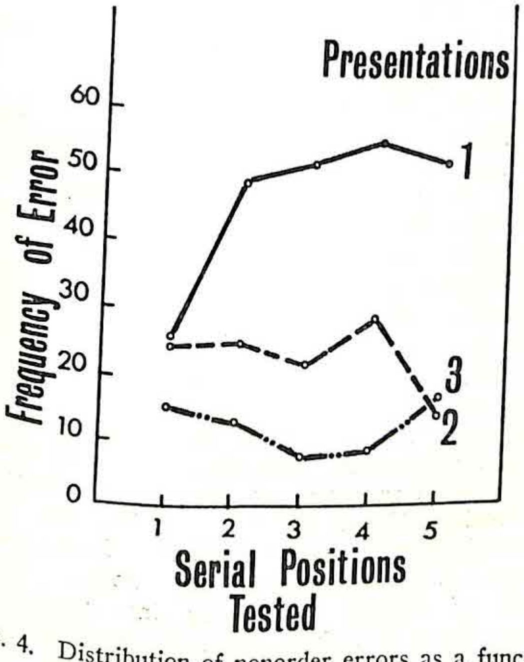


FIG. 4. Distribution of nonorder errors as a function of serial position tested.

TABLE 2

PERCENTAGE OF RESPONSES PER PRESENTATION FREQUENCY AMONG THREE CATEGORIES OF RESPONSES FOR "GOOD" AND "POOR" *S*S IN EACH GROUP

Presentation frequency	Good *S*s			Poor *S*s		
	1	2	3	1	2	3
4 sec.						
Correct R	61.3	77.3	88.0	30.7	58.7	68.0
Order errors	16.0	10.7	5.3	41.3	28.0	22.7
Other errors	22.7	12.0	6.7	28.0	13.3	9.3
8 sec.						
Correct R	36.0	73.3	81.3	18.7	45.3	53.3
Order errors	33.3	12.0	12.0	42.7	38.7	33.3
Other errors	30.7	14.7	6.7	38.7	16.0	13.3
16 sec.						
Correct R	41.3	72.0	80.0	22.7	32.0	45.3
Order errors	24.0	16.0	12.0	20.0	33.3	40.0
Other errors	34.7	12.0	8.0	57.3	34.7	14.7

Good and poor *S*s show comparable amounts of improvement with increased presentations at all retention intervals for both order and other kinds of errors.

The consistent differences between good and poor *S*s in correct responding suggest that repetitions serve the two groups differently; good *S*s, whose recall for content is initially high, learn order characteristics from repeated stimulus presentations, while poor *S*s must use repetitions to learn other characteristics of the words to permit content recall. Repetition helps good *S*s "tradeoff" order error for correct responses, while for poor *S*s the effect is to reduce serial position intrusions and omissions and increase content recall, which increases both correct responses and order errors. At 4 sec., e.g., the combined proportion of order error and correct responses for good *S*s constitutes 77.3%, 88.0%, and 93.3% of the responses at each presentation frequency. (For an order error to occur, a response from another serial position in the five-word string under test must occur, thus implying that content, but not order, recall is possible.) For the poor *S*s in the same condition, order errors and correct responses provide 72.0%, 86.7%, and 90.7% of the total responses. Clearly, at the short interval, both groups recall comparable amounts of content; good *S*s are characterized by a greater retention of order characteristics since the percentage of *correct* responses is consistently higher. The 8-sec. retention interval provides comparable data, while at 16 sec., poor *S*s show larger differences in correct recall and order errors, with considerable increase in other kinds of errors.

DISCUSSION

The pattern of errors observed in the present experiment supports the view that order and content errors can be separated experimentally. With a spatial cue for recall, "order errors" were distributed in a bowed serial position curve. Order errors occur when content is available but order information has been poorly learned or forgotten. This is most apparent in the data for "good" and "poor" *S*s; the total responses from within the five-word string were about equal for good and poor *S*s, but a greater proportion of these responses were given correctly (ie.., properly identified as having been in a particular serial position) by good *S*s than by poor *S*s, who made more order errors. The primary superiority of those with high proportions of correct responses over those with low proportions was *not* in the recall of content, but in the ordering of content correctly. Increased presentations helped poor *S*s to learn content and good *S*s to learn word sequences. The content and sequence of items in a series appear to be separable.

Errors based on nonorder or content loss have characteristics that are different from those of order errors. Nonorder errors (omissions, intrusions from previous lists, and extralist intrusions) appear to be invariant with serial position; such errors appeared most frequently during attempted recall of items from strings presented only once, and the intrusions from a previous list derived from strings presented two or three times. When content is unavailable from the string under test, omissions, extralist intrusions, or intrusions from a previous list will result. These errors fail to show a serial position effect and increase with longer retention intervals, while order errors show a clear serial position effect and show no significant increase with longer retention intervals. Intrusions from previous lists, while not order errors, show the effects of order. Order becomes the dimension that helps determine the subsequent occurrence of an intrusion in a subsequent list.

The equivalence of order errors, as defined here, to transposition errors, as defined by Conrad (1959, 1965) and Murdock and vom Saal (1967), is important to the evaluation of the role of order and content in recall errors. Murdock and vom Saal have shown that transposition errors occur quite infrequently but increase over retention intervals, and Conrad's (1959) data showed that good *S*s make more transposition errors than poor *S*s. The present data are contrary to both previous results. The discrepancy in findings may be attributed to specific differences in method or design or to the increase in content errors at 16 sec. in the present study which may have obscured order errors. The differences may also lie in the definition of order error. Explicit in Conrad's (1965) analysis is the assumption that transpositions are distributed according to the serial position effect. Dimensions of acoustic similarity may well account for transpositions in letter sequences; that this effect is more probable in the middle of these sequences suggests that such errors are facilitated by difficulty in ordering elements within the sequences. The problem of interpretation suggests that an analysis of transposition errors will provide only a partial answer to the problem of order error.

Order errors show a serial position effect. Order has been described as a dimension of the stimuli that, given the prescribed task, must be learned as part of a stimulus series. Elements of a series have an ordinal position; the learning of this characteristic of each element is improved with increased presentations of the stimulus sequence. In this view, order is a characteristic of sequentially organized stimuli that can serve as a dimension for classification and storage in memory. Turvey (1968) has pointed out in this regard that "the paucity of demonstrable dimensions of storage is only too apparent [p. 131]"; a dimension of similarity based on semantic differential measures (Turvey, 1968) and semantic relations among words (Henley, Noyes, & Deese, 1968) has been postulated to relieve the paucity. Serial order might usefully be considered another dimension of storage, particularly since generalization along an ordinal continuum has been suggested as a basis for the serial position effect

(Schulz, 1955) and has been discussed as a fruitful basis from which to explore the serial position effect in serial learning (Young, 1968). Indeed, order influenced the position of intrusion errors from previous lists in the present study. Table 1 shows that *S* frequently provided a response that had appeared in or adjacent to the position under test when, presumably, the correct response was unavailable.

The assumption that the identification of the ordinal position of an element in a sequence provides the basis for the serial position effect is found in Feigenbaum and Simon's (1962) model of the serial position effect in list learning. The first and last elements are assumed to be anchor points, with no prior or subsequent elements, respectively; it is assumed that these characteristics are learned independently of the remaining elements in the series, all of which have neighbors "before" and "after." Their unique status suggests, for the present data, that other words will be intruded in these positions with low probability and that words from these positions will intrude for other words in the sequence with low probability. The data of Fig. 2 support this interpretation for Positions 1 and 5 and show further that the intrusion pattern for words from the center of the list evidences a high degree of interdependence. The basis for the interdependence, defined by the higher intrusion rates after one presentation among Positions 2, 3, and 4, may well lie in generalization across these positions which, because they lack the unique status of the end positions, more readily interfere with, or are confused with, one another. Furthermore, Positions 2 and 4 evidence fewer order error intrusions from Positions 1 and 5, perhaps because of the relative uniqueness of these positions, and therefore show higher frequencies of correct recall than Position 3.

In terms of the present data, it seems clear that ordinal position of sequentially presented stimuli determines the kind and pattern of error. Errors related to content increase with retention interval and seem related to proactive inhibition effects in STM (Keppel & Underwood, 1962), but are unrelated to the serial position effect; errors or order may be unrelated to increases in retention interval, but are clearly related to the serial position effect. If *order* is an aspect of incoming stimuli, the functional stimulus for the spatial/temporal sequence is still to be determined, i.e., the nature of an order "tag" (Yntema & Trask, 1963) is still unknown.

REFERENCES

Brown, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

Conrad, R. Errors of immediate memory. *British Journal of Psychology*, 1959, **50**, 349–359.

Conrad, R. Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 161–169.

Craik, F. I. M. Types of error in free recall. *Psychonomic Science*, 1968, **10**, 353–354.

Crossman, E. R. F. W. Information and serial order in human immediate memory. In C. Cherry (Ed.), *Information theory*. London: Butterworth, 1960.

Feigenbaum, E. A., & Simon, H. A. A theory of the serial position effect. *British Journal of Psychology*, 1962, **53**, 307–320.

Fuchs, A. H., & Melton, A. W. Repetition and item length in short-term memory. *American Psychologist*, 1964, **19**, 730. (Abstract)

Henley, N. M., Noyes, H. L., & Deese, J. Semantic structure in short-term memory. *Journal of Experimental Psychology*, 1968, **77**, 587–592.

Keppel, G., & Underwood, B. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Melton, A. W., & von Lackum, W. J. Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition. *American Journal of Psychology*, 1941, **54**, 157–173.

Murdock, B. B., Jr. The criterion problem in short-term memory. *Journal of Experimental Psychology*, 1966, **72**, 317–324.

Murdock, B. B., Jr., & vom Saal, W. Transpositions in short-term memory. *Journal of Experimental Psychology*, 1967, **74**, 137–143.

Noyd, D. E. Proactive and intra-unit interference in short-term memory. Unpublished doctoral dissertation, University of Michigan, 1964.

Schulz, R. Generalization of serial position in rote serial learning. *Journal of Experimental Psychology*, 1955, **49**, 267–272.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, Bureau of Publications, 1944.

Turvey, M. T. Analysis of augmented recall in short-term memory following a shift in connotation. *British Journal of Psychology*, 1968, **59**, 131–137.

Yntema, D. B., & Trask, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

Young, R. K. Serial learning. In T. R. Dixon & D. L. Horton (Eds.) *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J:. Prentice-Hall, 1968.

(Received December 28, 1968)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 22–27

ATTENTIONAL PROCESSES ALONG A COMPOSITE STIMULUS CONTINUUM DURING FREE-OPERANT SUMMATION[1]

STANLEY J. WEISS[2]

American University

Predictions from a composite-stimulus-continuum analysis of summation were explored. Five rats were trained on a three-ply multiple schedule. Different variable-interval contingencies were associated with tone $(T + \overline{L})$ and light $(L + \overline{T})$, while simultaneous presentation of tone and light $(T + L)$ signaled extinction (S^{Δ}). During compounding tests, all *S*s summated to tone-plus-light absence $(\overline{T} + \overline{L})$, emitting more responses to this condition than to $(T + \overline{L})$ and $(L + \overline{T})$ conditions combined. These results establish the symmetry of the composite dimension postulated to underlie summation and indicate that the continuum extreme chosen as S^{Δ} determines that aspect of the composite discriminative stimulus controlling responding. They are used further to relate summation to generalization peak shift and to prove that stimulus intensity dynamism is not involved in multimodal summation.

Summation is observed when the simultaneous presentation of two or more *independently* conditioned stimuli leads to a greater response than any component stimulus presented separately. Although this phenomenon has been empirically established (see Weiss, 1963, 1964), customarily within a design that calls for the independent conditioning of stimuli from several modalities with extinction conditioned to the absence of these stimuli (Weiss, 1968), its dynamics have only been conjectured and remain obscure. The present experiment attempts to define the discriminal processes determining stimulus control along the composite stimulus dimension suggested to underlie summation (Wolf, 1963).

Weiss (1964) and Wolf (1963) performed two-modality, free-operant-summation studies that can, in principle, represent most multimodal summation studies to date, whether classical or operant, making their paradigms useful in describing the postulated composite continuum and its implications. Three conditions comprised their training: tone and light, each independently associated with food reinforcement, and the simultaneous absence of these two stimuli $(\overline{T} + \overline{L})$, signaling extinction ($S^{\Delta}$). This design is presented in the top row of Table 1. The composite continuum extends from light out, no tone L + T) through the individual presentation of tone $(T + \overline{L})$ or light $(L + \overline{T})$ to the tone-plus-light compound $(T + L)$. A binary analysis of the stimulus relations between the discriminative stimuli (S^Ds) and S^{Δ} indicates that each S^D contains an S^{Δ} element. But since summation was strongly demonstrated to $(T + L)$ in testing, it appears that the components of the $(T +)\overline{L}$ and $(L + \overline{T})$ S^Ds controlling responding were those S^D elements not also coincident with S^{Δ}, T, and L, respectively. Thus summation was forthcoming to the composite stimulus composed entirely of elements unique to an S^D function, which, lacking discriminated S^{Δ} elements, was further removed from S^{Δ} on the postulated continuum than either of the other test stimuli.

The composite continuum is defined by the on–off states of all relevant discriminative or conditioned stimuli, one continuum extreme anchored to the all-off state, the other to all on. The points between these continuum extremes are scaled with refer-

---

[1] This research was supported by Grants MH-13495 and MH-16853 from the National Institute of Mental Health, United States Public Health Service. The author would like to express his appreciation to John Monahan and Judith Roes for their capable assistance in running *S*s. This study was presented at the meeting of the Eastern Psychological Association, Philadelphia, Pennsylvania, April 1969.

[2] Requests for reprints should be sent to Stanley J. Weiss, Department of Psychology, American University, Washington, D. C. 20016.

TABLE 1

SCHEMATIC REPRESENTATION OF TRAINING AND TEST CONDITIONS OF PREVIOUS AND PRESENT TWO-MODALITY FREE-OPERANT-SUMMATION DESIGNS

Composite stimulus continuum	Light out, no tone $(\overline{T}+\overline{L})$	Tone $(T+\overline{L})$	or	Light $(L+\overline{T})$	Tone and Light $(T+L)$
Design employed in previous two-modality free-operant-summation studies[a]	Training				
	S^{Δ} [b]	S_1^D [c]		S_2^D	Summation obtained
		Test			
Design of present experiment		Training			
	Summation predicted	S_1^D		S_2^D	S^{Δ}
	Test				

[a] Wolf, 1963; Weiss, 1964, 1968.
[b] S^{Δ} = stimulus condition associated with extinction.
[c] S^D = stimulus associated with reinforcement.

ence to the number of elements common to each. In the two-element continuum employed by Weiss (1964) and Wolf (1963), the continuum anchors would be $(\overline{T}+\overline{L})$ and $(T+L)$, with the individual S^Ds containing one element of each anchor. Theirs was therefore a three-step composite continuum, the number of points on any composite scale being equal to the number of independently conditioned S^Ds, or CSs, plus one.

The present experiment tested the symmetry of the two-element continuum described for the Weiss (1964) and Wolf (1963) studies. When the locus of S^{Δ} is reversed to the $(T+L)$ continuum extreme, with the S^Ds maintained as $(T+\overline{L})$ and $(L+\overline{T})$, the elements of each S^D controlling responding should be $\overline{L}$ and $\overline{T}$, respectively. This derives from a mirror-image analysis of those attentional processes operating when S^{Δ} is $(\overline{T}+\overline{L})$. Therefore, that composite condition composed entirely of elements unique to an S^D function, thereby lacking S^{Δ} elements and thus furthest removed from S^{Δ} on the composite scale, is $(\overline{T}+\overline{L})$, or "stimulus absence." According to the composite-continuum analysis of stimulus control operating, summation should occur to $(\overline{T}+\overline{L})$. The design of the present experiment, presented in the bottom row of Table 1, tested this prediction, which, if upheld, would reinforce the existence of the composite continuum postulated as well as its symmetry.

METHOD

Subjects.—Five naive adult male hooded rats, approximately 350 gm. at the start of deprivation, were trained and tested at 80% of their free feeding weight.

Apparatus.—The operant training chamber measured 8 in. high, $8\frac{3}{8}$ in. long, and 7 in. wide. Its front and rear walls were aluminum, its side walls $\frac{1}{4}$-in. white translucent plastic, and its floor and ceiling $\frac{1}{2}$-in. hardware cloth. A Gerbrands microswitch bar, requiring approximately 15 gm. for depression, was located on the right side of the front wall, $2\frac{3}{4}$ in. above the floor. A feeder trough was on the left side of the front wall at floor level. An 8-in. Quam 8C12X-8 speaker was in an enclosure centered $8\frac{1}{4}$ in. above the training chamber. A 6-in., 25-w., 120-v. General Electric showcase bulb was horizontally mounted 4 in. from the middle of each of the translucent side walls. These were enclosed in specially ventilated clear plastic cylinders to augment heat dissipation. The training chamber and its associated stimulus presentation devices were enclosed in a sound attenuation chamber described elsewhere (Weiss, in press). The tones used were inaudible outside this attenuation chest.

The 2,000-cps, approximately 95-db. tone employed in training and testing was generated by a BRS AA-201 audio oscillator and amplified by a BRS AO-201 amplifier. The sound level measurement was taken with a General Radio Type 1565-A sound level meter, Scale C, with the microphone located on the ceiling of the training chamber parallel to the speaker. The ambient noise level, with the exhaust fan on, was approximately 80 db.

The S^D light intensity employed in training and testing was approximately 38 ftc. measured with a Honeywell Pentax 1°/21° photometer held 5 in. from, and aimed at, the chamber's side walls, with both showcase bulbs operating on 115 v. The training chamber was very dimly illuminated at all times by a shielded 7-w., 120-v. bulb operating at 3 w. The houselight was too dim to activate the photometer, but allowed *E* to see *S* within the apparatus.

Solid-state programming equipment was located in a room adjacent to that housing the training chamber. Reinforcers were Noyes 45-mg. rat pellets.

Procedure.—After magazine training and shaping, the terminal training contingency was gradually approximated. The three-ply multiple schedule contained discriminative stimulus (S^D), associated variable-interval (VI) 30-sec. and VI 70-sec. components, plus an extinction (S^Δ) period. The S^Ds, averaging 3.5 min. but varying between presentations within the limits of 2.5 and 4.5 min., were counterbalanced tone $(T + \bar{L})$ and light $(L + \bar{T})$. The operant was bar pressing. When a VI schedule was in effect, a reinforcement was produced by the first response that occurred an average of 30 or 70 sec. after the previous reinforcement, depending on which schedule was in effect. The limits of the VI 30-sec. schedule were 2 and 80 sec., of the VI 70-sec. schedule, 2 and 180 sec. These intervals were sequenced to keep the time until reinforcement availability, during any interval, independent of the preceding interval.

The termination of each S^D-associated reinforcement component of the training schedule was followed by an S^Δ period correlated with the simultaneous presentation of tone and light $(T + L)$. This S^Δ required the cessation of bar pressing for a minimum time for one of the VI S^Ds to reappear. The no-response requirement was varied, within sessions, within the limits of 20 and 80 sec. The *S*s were placed in the chamber with the S^Δ stimulus on. When a component change occurred, it was correlated with *only* the addition or removal of the appropriate stimulus. The other stimulus remained *undisturbed* during this transition. *S*s 17, 22, and 28 had $(L + \bar{T})$ associated with their VI 30-sec. schedules and $(T + \bar{L})$ with their VI 70-sec. schedules. These stimulus–schedule combinations were reversed for *S*s 26 and 31.

All *S*s were trained for approximately 30 sessions on the schedule described. Each session lasted 3 hr. or until 130 pellets were received, whichever occurred first. After the training, they were tested for summation. The extinction test consisted of presentations of the tone $(T + \bar{L})$, the light $(L + \bar{T})$, and the absence of these stimuli—light out, no tone $(\bar{L} + \bar{T})$. The stimulus conditions were presented in each of 18 block randomizations. Each test condition was presented 45 sec. per block with a 15-sec. $(T + L)$ period (formerly S^Δ) separating test conditions. Testing started after *S* had received approximately 50 pellets on his training schedule.

RESULTS

Figure 1 presents the terminal training performance of two *S*s, each representing one of the stimulus–schedule combinations used in training. These cumulative records clearly show good differentiation between VI 30- and VI 70-sec. S^Ds, with the differential rates appropriate to each component of the multiple schedule *commencing at the onset* of the respective stimulus presentations. Table 2 presents this terminal performance data quantitatively. All *S*s responded at a higher rate during the VI 30-sec. than the VI 70-sec. S^D, with clear rate distinction between S^D and S^Δ conditions.

Table 3 shows the total response outputs to each test condition. Excluding that condition to which *S*s were taught to cease responding $(T + L)$ so as not to overemphasize the overall significance of the analysis of variance, a Treatment × *S* analysis of the $(T + \bar{L})$, $(L + \bar{T})$, and $(\bar{T} + \bar{L})$ test conditions yielded an $F\ (2,\ 8) = 7.00$, $p < .025$. Significantly more responses ($p < .025$) were emitted to the $(\bar{T} + \bar{L})$ condition than to either the VI 30- or VI 70-sec. stimuli, $(T + \bar{L})$ or $(L + \bar{T})$, separately presented, with the difference between VI 30-sec and VI 70-sec. stimuli insignificant. Each *S* emitted more responses to the $(\bar{T} + \bar{L})$ test condition than he emitted to the separate $(T + \bar{L})$ and $(L + \bar{T})$ conditions combined. Dividing the total responses emitted to the $(\bar{T} + \bar{L})$ test condition by the sum of the responses emitted to the tone and the light stimuli individually presented yields ratios $[(\bar{T} + \bar{L})/(T + \bar{L}) + (L + \bar{T})]$ of 1.4, 1.6, 1.6, 1.9, and 2.5 for the five *S*s, respectively.

The progress of summation over test replications is seen in Fig. 2. This figure presents the mean cumulative percentage of responses emitted to each test condition

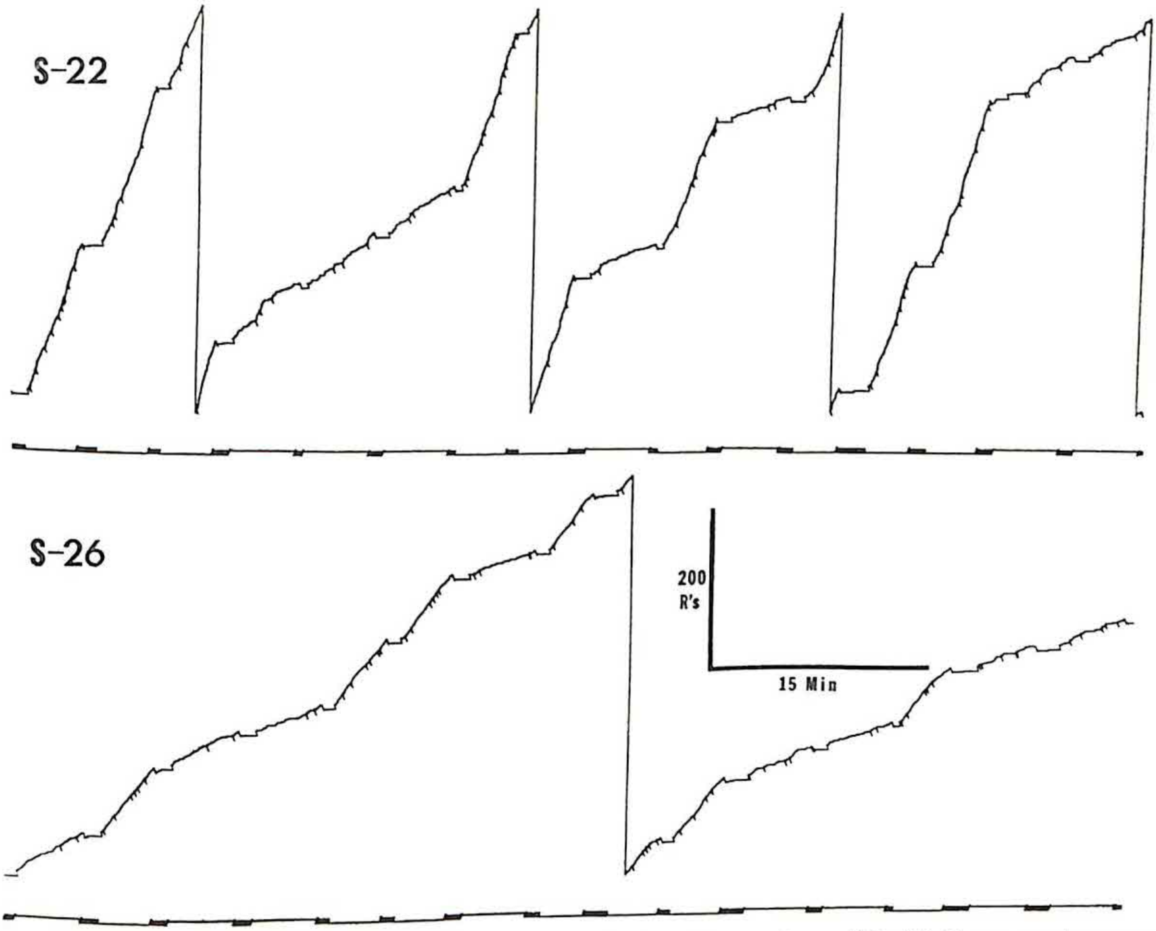


FIG. 1. Cumulative records of *S*s 22 and 26 from a terminal training session. (The VI 30-sec. contingency is effective when the base line is in the lower register, and the VI 70-sec. when it is elevated. The solid portions of the base line and depression of the cumulative response pen identify the S^{Δ} tone-plus-light periods. Slash marks by the latter pen record reinforcements. For *S* 26, $(T + \bar{L})$ was paired with the VI 30-sec. schedule and $(L + \bar{T})$ with the VI 70-sec. schedule; the opposite stimulus–schedule combinations were employed for *S* 22.)

over replications. It is clear that summation to the light-out, no-tone condition occurred from the very beginning of testing and was maintained through the final test replications. The *S*s continued to emit responses to $(\bar{T} + \bar{L})$ when responding to presentations of $(T + \bar{L})$ or $(L + \bar{T})$ was essentially extinguished.

DISCUSSION

The postulation of a symmetrical composite stimulus dimension in summation designs has been supported. Summation will occur to the stimulus complex representing either extreme of a two-modality composite continuum when "single" stimulus presentations are reinforcement-correlated if the opposite extreme is conditioned as S^{Δ}. Therefore the continuum is symmetrical, with the functions of the two extremes interchangeable.

Schematically, the stimulus complexes that are reinforcement-associated are $(T + \bar{L})$ and $(L + \bar{T})$, irrespective of whether S^{Δ} is $(T + L)$, as in the present experiment, or $(\bar{T} + \bar{L})$, as in the Weiss (1964) and Wolf (1963) studies. But the summation tests clearly show that when S^{Δ} is $(\bar{T} + \bar{L})$, the T and L elements of these composite S^{Δ}s are controlling responding, while when S^{Δ} is $(T + L)$, the $\bar{L}$ and $\bar{T}$ elements are serving this function. Therefore, the composition of S^{Δ} relative to the S^{D}s serves to determine those elements of the

TABLE 2

Terminal Training Performance (Responses per Minute)

S	Schedule condition		
	VI 30 $(T + \bar{L})$ or $(L + \bar{T})$	VI 70 $(L + \bar{T})$ or $(T + \bar{L})$	S^Δ $(T + L)$
17	5.2	4.6	.4
22	58.8	16.1	1.6
26	21.8	8.7	1.3
28	12.1	7.7	.5
31	25.5	11.2	2.5
M	24.7	9.6	1.3

Note.—The rates represent averages of the final three training sessions for each *S*. The initial half hour of each session was considered warm-up and omitted from the determinations.

composite S^D to which *S* is selectively attending. Observing *S* responding during $(T + \bar{L})$, or $(L + \bar{T})$ alone does not by itself allow this conclusion. It is the compounding test that permits isolation of the control exerted by the two elements of these composite S^Ds.

Weiss (1964) reported a compounding study that employed training conditions essentially similar to the present experiment with the exception that S^Δ was $(\bar{T} + \bar{L})$ rather than $(T + L)$. The training cumulative records presented in that study (Weiss, 1964, Fig. 1) show stimulus control that compares favorably with that demonstrated by *S*s of the present experiment in Fig. 1. Differentiation between the S^Ds and S^Δ is clear in both studies, as well as different rates to the two S^Ds. Although the distribution of responses over the test conditions of the present study cannot be directly compared to that of Weiss (1964) because of differences in testing procedures, the comparison can be made with data reported by Weiss (1968). Four *S*s from the latter study were given essentially similar training and testing as that received by *S*s in the present experiment with the exception that S^Δ was $(\bar{T} + \bar{L})$. (The apparatus and stimuli employed by Weiss, 1968, are described in Weiss, 1967.) The mean percentages of responses emitted to each condition over the first 18 test replications of the Weiss (1968) study were: S^D composite $(T + L)$, 55.1; VI 30 S^D, 21.9; VI 70 S^D, 19.5; and $S^\Delta(\bar{T} + \bar{L})$, 3.5. Figure 2 shows the comparable distribution of responses in the present experiment. This distribution is, in percentages: S^D composite $(\bar{T} + \bar{L})$, 60.8; VI 30 S^D, 19.5; VI 70 S^D, 15.8; and S^Δ $(T + L)$, 4.0. The greatest difference between comparable composite conditions of the two studies is 5.7%, the least, .5%. The essentially similar distribution of test responses, as well as the similarity of the training records discussed, supports the conclusion that the behavioral control to comparable aspects of the composite continuum is equivalent even though opposite extremes serve as S^Δ and different elements of the composite S^Ds control responding.

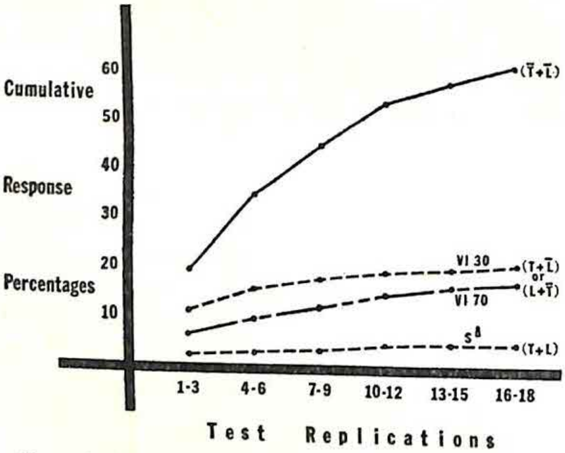


FIG. 2. Mean cumulative response percentages emitted to each test condition over replications. (The stimulus condition represented by each curve is indicated by the notation to the extreme right. $(\bar{T} + \bar{L})$ indicates the light-out, no-tone condition. The presentation of $(T + \bar{L})$ or $(L + \bar{T})$, as appropriate, is represented by the VI 30-sec. and VI 70-sec. curves. S^Δ represents the tone-plus-light $(T + L)$ periods.)

TABLE 3

Summation Test Responses

S	Stimulus condition			
	$(\bar{T} + \bar{L})$	VI 30 $(T + \bar{L})$ or $(L + \bar{T})$	VI 70 $(L + \bar{T})$ or $(T + \bar{L})$	S^Δ $(T + L)$
17	86	28	32	4
22	640	160	101	24
26	197	81	40	15
28	165	41	56	10
31	347	135	86	36
M	287	89	63	17.8

The symmetry of the control supports Wolf's (1963) comparison of the response enhancement effect of combined S^Ds to generalization peak shift, with the composite stimulus dimension employed in summation experiments functionally equivalent to the unitary dimension of the peak shift design. Peak shift occurs during stimulus generalization tests after discrimination training in which one value on a continuum is reinforced and another extinguished (Hanson, 1959). These studies have shown that either value along the stimulus continuum can serve equally well in either S^D or S^Δ capacity, with the peak shifted one way or the other as a function of the position of S^Δ relative to S^D. The reversibility of function has now also been demonstrated in summation, lending support to Wolf's suggestion.

From a physical point of view, summation has been demonstrated in the present study to the *least intense* test stimulus ($\bar{T} + \bar{L}$). Therefore, these results have significant implications for the application of a stimulus-intensity-dynamism interpretation to summation. Simply postulated, this interpretation might contend that the compound condition (of previous summation studies) was more intense than each of the "single" stimuli comprising it, and stimulus intensity according to Hull (1952) is an energizer of behavior. Since in the present experiment summation was as *powerfully* demonstrated to ($\bar{T} + \bar{L}$) as it has previously been reported to (T + L) in comparable "mirror-image" compounding studies, it appears that stimulus intensity dynamism is in *no way* responsible for summation to a multimodal stimulus. Rather, the present study shows that summation is a function of the relationship between S^D and S^Δ stimulus conditions on a composite continuum. Although the simplest composite continuum was investigated in the present study, more profound predictions are possible when three- or even four-element composites are explored. This suggests that the stimulus compounding procedure is a valuable assay technique for the investigation of the determinants of selective attention and stimulus control.

REFERENCES

Hanson, H. M. Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, 1959, **58**, 321–334.

Hull, C. L. *A behavior system.* New Haven: Yale University Press, 1952.

Weiss, S. J. The summation of response strengths instrumentally conditioned to stimuli along different sensory modalities. Unpublished doctoral dissertation, Ohio State University, 1963.

Weiss, S. J. Summation of response strengths instrumentally conditioned to stimuli in different sensory modalities. *Journal of Experimental Psychology*, 1964, **68**, 151–155.

Weiss, S. J. Free-operant compounding of variable-interval and low-rate discriminative stimuli. *Journal of the Experimental Analysis of Behavior*, 1967, **10**, 535–540.

Weiss, S. J. Response inhibition as a factor in additive summation. Paper presented at the meeting of the Eastern Psychological Association, Washington, D. C., April 1968.

Weiss, S. J. Design of an effective economical sound attenuation chamber. *Journal of the Experimental Analysis of Behavior*, in press.

Wolf, M. M. Some effects of combined S^Ds. *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 343–347.

(Received January 4, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 28–33

PERMUTATION TRANSFORMATIONS ON PHRASE STRUCTURES IN LETTER SEQUENCES[1]

TERRENCE J. KEENEY[2]
University of Minnesota

A paired-associate (PA) learning task in which the stimuli were strings of five consonants followed by a numeral and the responses were reorderings of the same five consonants was presented to college *S*s. Each response string was related to its stimulus string by one of four permutation rules. Each rule was associated with one of the stimulus numerals. The four rules together segmented the consonant string into a hierarchical system of units. With standard PA instructions, there was an average of 4.06 trials to criterion in initial learning. When subsequently transferred to a list with five entirely new consonants but the same rules, *S*s performed virtually errorlessly. In responding, *S*s revealed a tendency to pause in a pattern appropriate to the hierarchical units imposed by the permutation rules. However, the similarity of these experimentally induced rules with those involved in natural language behavior was rendered doubtful by the long experimental response latencies of nearly 5 sec. per rule.

What is learned by a child when he acquires the ability to speak and to comprehend a natural language has been a matter of dispute. The knowledge that a speaker possesses of his language has been viewed, on the one hand, as a complex network of stimulus-response connections (Jenkins & Palermo, 1964; Mowrer, 1960; Osgood, 1953) and, on the other hand, as a system of rules which generate the well-formed utterances in his language (Chomsky, 1957, 1965). According to the latter view, phrase structure rules and transformational rules are essential features of the syntax of all natural languages. Each of these types of rules are general enough that they can be applied to many different materials in addition to natural language. For example, Miller (1967) and Shaw and Pufall (1968) have studied two-choice behavior within a rule-learning context.

For the present investigation, a paired-associate learning task was devised in which the stimuli could be analyzed according to phrase structure and transformational rules. These rules governed the relationship between the members of each item pair. The learning of the pairs was analyzed in order to determine if it involved the formation of such rules.

EXPERIMENT I

A paired-associate learning task in which there were three distinct elements in each "pair" was presented visually to *S*s. The left-hand portion of the display consisted of a five-consonant string; the upper central portion of the display consisted of a single numeral; and the right-hand portion of the display consisted of another string of five consonants. The consonants on the right were identical with those on the left; only their order differed from one side of the display to the other.

A set of phrase structure rules was imposed on each consonant string. Phrase structure rules are rules which simply expand one symbol into one or more other symbols. These rules had the effect of segmenting the five elements of each string into

---

[1] This study was based on a dissertation submitted to the faculty of the University of Minnesota in partial fulfillment of the requirements for the PhD degree. The research was carried out while the author was a predoctoral fellow at the Center for Research in Human Learning, University of Minnesota, and was supported in part by grants from the National Science Foundation, the National Institute of Child Health and Human Development, and the Graduate School of the University of Minnesota to the Center. This article was written while the author was at Educational Testing Service, Princeton, New Jersey.

[2] Requests for reprints should be sent to Terrence J. Keeney, who is now at the University of California, Department of Psychology, Riverside, California 92502.

TABLE 1

Phrase Structure and Permutation Rules Applied to a Five-Element String

Phrase structure	Phrase structure rule	Phrase marker	Permutation rule			
			1	2	3	4
PS$_1$	4 → A + 3 3 → 2 + 1 2 → B + C 1 → D + E	((A)((BC)(DE)))	((A)((BC)(ED)))	((A)((CB)(DE)))	((A)((DE)(BC)))	(((BC)(DE))(A))
PS$_2$	4 → 3 + E 3 → 2 + 1 2 → A + B 1 → C + D	(((AB)(CD))(E))	(((AB)(DC))(E))	(((BA)(CD))(E))	(((CD)(AB))(E))	((E)((AB)(CD)))
PS$_3$	4 → 3 + 1 3 → A + 2 2 → B + C 1 → D + E	(((A)(BC))(DE))	(((A)(BC))(ED))	(((A)(CB))(DE))	(((BC)(A))(DE))	((DE)((A)(BC)))

a system of hierarchical units (i.e., a phrase marker). These rules were formally identical to the phrase structure rules employed in the description of natural languages. The three sets of phrase structure rules employed in these experiments and the phrase markers they define are presented in Table 1.

The relationship between the left- and right-hand strings of each pair was defined by a permutation transformational rule. These permutation rules reversed the order of two adjacent units as defined by the phrase structure rules. Since these units could consist of as many as four consonants or as few as one consonant, permutation rules rearranged a variable number of consonants, depending on the particular rule being employed. The four permutation rules as they operate on each of the three phrase markers are also presented in Table 1.

Method

Design.—Six different treatment groups were used in the present study. Groups 1, 2, and 5 initially learned the materials constructed according to Phrase Structure 1 (PS$_1$), while Groups 3, 4, and 6 initially learned the materials from Phrase Structure 2 (PS$_2$). In order to determine what was learned in the course of mastering this first list, a transfer task was employed. If learning the initial task consisted solely in the formation of associations between the left- and right-hand strings, then the introduction of five entirely new consonants should have completely disrupted performance. If, on the other hand, during the course of initial learning, permutation rules were learned, then changing the letters should have had little effect on performance. Groups 1, 2, 3, and 4 were all subjected to a complete change in letters from the first to the second task. All of these groups were presented with the J letter set (J,X,C,P,M) in initial learning and then with the N letter set (N,S,K,B,Q) in the transfer task. Groups 5 and 6 were presented with the N letter set on their initial, and only, task in order to control for the possible differential effect of the letter sets on learning speed.

If the initial learning were solely on the basis of associative connections between the letter strings, then transfer performance for all of the groups should have been uniformly very poor. In order to determine if, in addition, rules were being learned, Groups 2 and 4 were presented with strings constructed according to a new set of phrase structure rules in the transfer task. Both rule learning and the formation of specific associations in initial learning for these groups would interfere in transfer performance; warm-up and nonspecific transfer effects would be the only facilitating factors. Since initial learning for Group 2 was on PS$_1$, the transfer list was based on PS$_2$, while the situation was exactly the opposite for Group 4. If initial learning were on the basis of rules, then Groups 1 and 3 would have been expected to perform very well on the transfer task and better than Groups 2 and 4.

Subjects.—The *S*s were 54 undergraduate female volunteers from the University of Minnesota, divided into six groups with nine *S*s per group. They were randomly assigned to one of the six groups according to their order of appearance at the laboratory.

Materials.—The consonants in the two letter sets were chosen on the basis of their low intraset visual and auditory confusability. The mean rank frequency of occurrence in English (15.6) was identical for the N set and the J set (Bourne & Ford, 1961), while the average meaningfulness for the letters in the N set was 10.9 and for the J set, 11.2 (Anderson, 1965). None of the primary or secondary responses to any of the letters was present within a set, but four such associations occurred between the sets (N and P, M and N, J and K, P and Q) (Underwood & Schulz, 1960). These associations potentially could have aided transfer performance, but in only 26% of the transfer instances were these associations appropriate and, consequently, even potentially facilitating.

Three random orderings of the five consonants in both the N set and J set were prepared. These strings served as the left-hand portions of the display. The four permutation rules appropriate to both of the phrase structures were then applied to each of these three left-hand strings to yield the right-hand strings. Since there were four different permutation rules for each phrase structure and three different left-hand strings within each letter set, 12 consonant strings from the N set and 12 from the J set were generated for each phrase structure. The center portion of the display consisted of the numerals "1 2 3 4," with one of the numerals underlined. Each underlined numeral was arbitrarily assigned to a permutation rule and served as the cue for application of this rule to the left-hand string. Each of the three portions of the display was placed on a separate slide for projection onto a screen.

Procedure.—The stimulus in both tasks consisted of the left-hand string and an underlined numeral. The *S*s were required to recite from left to right the letters that were paired with this string-numeral combination. They were not told what the five letters on the right would be or, of course, their sequential order. The left-hand string was not bracketed in any way, nor was any other indication of the phrase structure of the strings ever given to *S*s. No mention of permutations, rearrangements, or of rules of any sort was ever made. The *S*s were asked to recite the answer as rapidly as possible after the left-hand letters and the numeral were flashed on the screen. After giving the response verbally, *S*s were required to press a button placed at the base of the screen, which response caused the correct answer to be shown on the screen. Guessing was encouraged.

There was no time requirement placed on response latency, except for the instruction to respond as rapidly as possible. After the response, all three portions of the display remained on for 9.5 sec., during which time *S*s were able to compare the left- and right-hand letter strings and to note the underlined number. At the termination of this feedback period, all three portions of the display went blank for 1 sec. until the next stimulus appeared. No verbal feedback was given.

With each of the four permutation rules applied to each of the three left-hand strings, a complete trial consisted of 12 item pairs. These item pairs were arranged in three random orders, with the following restrictions: First, the first four items in each order contained one instance of each rule; second, two or more instances of the same rule were never presented consecutively. These three random orders were presented consecutively to each *S*, with the starting point (at the beginning of Order 1, 2, or 3) assigned randomly for each *S*. A 10-sec. intertrial interval was employed.

The learning criterion on both tasks was two successive errorless trials. If an *S* did not learn the list within nine trials, the experiment was terminated. Immediately after *S* reached criterion, he was transferred to the appropriate second task condition, with no further instructions by *E* and no longer than the usual 10-sec. intertrial interval.

In light of the evidence presented by Martin and Strange (1968) that in natural speech when *S*s have little uncertainty as to message content, they tend to pause at phrase boundaries, the pausal patterning in the experimental responses was analyzed. Twenty-six volunteers from the introductory psychology course at the University of Minnesota listened to the recorded responses of the last two trials (the errorless trials) of first-list learning. A total of 22 *S*s' responses was analyzed for pauses. Eleven of these *S*s had learned PS_1 and 11 had learned PS_2. Within each phrase structure, *S*s were chosen randomly. The errorless trials were chosen for analysis to minimize the chance that there would be instances of a hesitation due to uncertainty before an incorrect response. Only first-list responses were included in order to avoid the possible residual effects of a structure learned earlier in the experiment on subsequent responding.

Three groups of judges, two groups of 7 and one group of 12, listened to the recorded responses and judged them as belonging to either a "da—da da—da da" or a "da da—da da—da" pattern. This forced-choice technique was valid since these were the only two patterns imposed by the two phrase structures used. Rules 1, 2, and 3 of PS_1 and Rule 4 of PS_2 yield the I-II-II pattern. Rule 4 of PS_1 and Rules 1, 2, and 3 of PS_2 yield the II-II-I pattern. Each group of judges evaluated an equal number of utterances from each phrase structure. The two groups of 7 judges judged four *S*s' responses from each of the phrase structures (eight *S*s in all), while the group of 12 judges judged three *S*s' responses for each of the phrase structures (six *S*s in all).

The judges were instructed to categorize each response as belonging to one of the two pausal patterns. Prior to the experimental task, *E* gave an example of each pattern. A judgment had to be made on each response.

Results

The mean number of trials to criterion on first-list learning for each treatment is given in Table 2. Only 2 *S*s did not reach criterion on the first list within the experimental session, 1 each from Group 2 and Group 3. These 2 *S*s were replaced by new *S*s and their data were discarded. Thus, 54 out of the total of 56 *S*s presented with the initial learning task were able to reach

TABLE 2

Trials to Criterion for All Groups, First and Second Lists, Exp. I

Item	Group					
	1	2	3	4	5	6
First list						
PS	1	1	2	2	1	2
$\bar{X}$	4.44	4.00	3.89	3.89	3.33	4.78
SD	1.59	1.87	1.17	1.54	1.32	1.56
Second list						
PS	1	2	2	1		
$\bar{X}$	.78	2.67	.44	3.11		
SD	1.39	1.22	.73	1.54		

Note.—PS = phrase structure.

within nine trials the criterion of two successive correct trials. In first-list learning, the effect of letter set (Groups 5 and 6 vs. Groups 1, 2, 3, and 4) was not significant.

The mean number of trials to criterion during second-list learning for each treatment group is also presented in Table 2. All 36 of the *S*s who were given the transfer task reached criterion within the allotted nine trials. The data from the two groups that transferred within the same phrase structure (Groups 1 and 3) were combined and contrasted with the data from the two groups that transferred to a new phrase structure (Groups 2 and 4). A difference score (List 1 − List 2) was computed for each *S*. The mean difference for Groups 1 and 3 was significantly greater than the mean difference for Groups 2 and 4, $t\ (34) = 4.46$, $p < .001$.

The responses that were being judged as to their pausal patterns were either from Phrase Structure 1 or 2. Combining the data from all of the judges and *S*s, 68% of the judgments of responses in PS_1 were correct and 60% of the judgments of responses in PS_2 were correct. The base line, which would indicate chance performance, for the previous percentages is 50%. Thus the 68% and 60% correct judgments for Phrase Structures 1 and 2, respectively, were not impressively large. However, when a tabulation of the correct judgments was made, summed over the 7 or 12 judges who listened to each *S*, 9 of the 11 *S*s in PS_1 and 8 of the 11 *S*s in PS_2 had more of their responses judged correctly than incorrectly. An analysis of these data with Yates' correction for noncontinuity yielded a $\chi^2\ (1) = 4.58$, $p < .05$.

Discussion

The results of Exp. I indicate that the experimental task is one which can be mastered with little difficulty by college sophomores. It is apparent from the results on the second task that in learning the first list, *S*s inferred some rules which they could subsequently apply to an entirely new string of letters.

The small but significant tendency, which the judges in the present study exhibited, toward correct identification of the pausal patterning in the responses of *S*s in Exp. I supports the conclusion that in the course of learning, *S*s incorporated a psychological analogue to the formal phrase structure. Having never seen the phrase structure rules, *S*s, who had learned only transformational rules which were compatible with the phrase structure, transmitted information about the nature of the phrase structure in giving their responses. A number of factors were militating against the transmission of such pausal patterns in responding. The emphasis on speed in giving the response and the short length of the response strings (making it unnecessary to pause in the midst of responding) were two such factors. Given a more leisurely pace and longer strings of symbols, the pausal patterns might well have been more pronounced.

Experiment II

Experiment I dealt with the operation of only single permutation rules. However, one feature of permutation rules is that more than one such rule can be applied to a given string and a uniquely determined string still emerges as a result. Once each of the four rules has been learned in its singular operation, there is no formal reason why there should be any decrement in performance when two or three of the rules are applied simultaneously to the same string. Given the human cognitive apparatus, however, performance factors such as memory limitations may cause disruption of performance in the multiple application of these rules. Furthermore, if the rules, once learned, become cognitive operations and if these operations are performed sequentially, then latency measures on the responses should increase as the number of rules applied to a string increases. Experiment II investigated the performance of single, double, and triple rule operations on the strings.

Method

Subjects.—Nine male undergraduates from the University of Minnesota served as *S*s.

Materials.—The set of phrase structure and transformational rules defined by Phrase Structure 3 (PS_3) (see Table 1) were imposed on the consonant strings. The four individual permutation rules could be combined into six double rules (i.e., 1 and 2, 2 and 3, etc.) and four triple rules (i.e., 1 and 2 and 3, 1 and 3 and 4, etc.). The order of application of the rules was arbitrarily fixed so that the rules were applied in accordance with their numerical order. Thus, Rule 1 was always applied before Rule 2, Rule 2 before Rule 3, etc. This was a purely arbitrary, but necessary, convention to ensure that all *S*s would be applying the rules in the same order and thus produce the same response.

Instead of numerals, pictures of the four suits in a deck of cards were used as cues for the rules. Slides with one, two, or three symbols appearing one beside the other were prepared for projection onto the central portion of the display. The suit symbols were randomly assigned to the rules and displayed so that a left-to-right ordering of the symbols conformed to the proper ordering of the rules. A total of 14 different combinations of the symbols was constructed. An equal number of separate random orderings of the consonants was also constructed to be employed as the left-hand string.

Procedure.—Initial learning of the singular rules was facilitated by giving *S*s a 5-min. period of free learning in which they were allowed to study one example of the singular application of each of the four rules. Prior to this free learning, *S*s were informed that at the end of the 5 min. they would be required to anticipate the right-hand string when shown the left-hand string and the suit symbol.

Following the free-learning period, *S*s were presented with a task which was procedurally identical to the learning task in Exp. I. Three random orders of the N letter set were presented as the left-hand strings; the four suit symbols were presented one at a time as the central portion of the display; and the strings appropriate to PS_3 were presented as the right-hand strings. Once the criterion of two successive correct trials was met, *S*s were transferred to the multiple-rule transfer list. They were instructed to apply the rules one by one, starting with the rule cued by the leftmost symbol. The string resulting from the application of this first rule was then to be used as the string to which the next rule was to be applied, etc., until the final "answer" was computed. The *S*s were instructed to give only the final answer aloud and to give it as rapidly as possible.

TABLE 3

Mean Number of Errors, Percentages of the Total Opportunity for Errors, and Mean Latencies (Trials 2–5), Exp. II

Item	Number of rules applied		
	1	2	3
Mean errors	0	1.11	2.22
Percentage of errors	0%	3.7%	11.1%
Mean latency (in sec.)	5.07	9.09	15.34

Five trials on this 14-item list were presented to each *S*. Two randomizations of the transfer list were constructed. All *S*s were exposed to these two randomizations in alternating order, resulting in three exposures to the first randomization and two exposures to the second randomization. Latencies were recorded for all of the responses.

Results

Following the 5 min. of free learning, *S*s mastered the task rapidly. The mean number of trials to criterion was 1.89 ($SD = .71$). The mean number of errors was 4.90, with a range from 1 to 18.

For single, double, and triple applications of the rules, the mean number of errors and the percentage of errors out of the total opportunities for errors are shown in Table 3. The percentage of errors was quite small in all of the cells, the largest being less than 15%, indicating that *S*s performed on even the triple rules with a high degree of accuracy. The number of errors increased consistently with the increasing number of rules applied. A Friedman two-way analysis of variance revealed a significant difference in the number of errors committed across the three levels of rule application, $\chi^2_r\,(2) = 12, p < .01$.

The mean latencies over Trials 2–5 of the transfer test for single, double, and triple applications of the rules for the correct responses only are shown in Table 3. The first transfer trial and the incorrect responses were excluded from these data because of the extraordinarily long latencies exhibited by many of the *S*s on their first few encounters with a multiple rule and on their incorrect responses. Within either single, double, or triple applications of the rules, there was no significant difference among the particular rules in the number of errors committed. A one-way analysis of variance on the mean latencies revealed the effect of number of rules to be significant, $F\ (2, 16) = 30.36$, $p < .01$. A trend analysis performed on the number-of-rules factor yielded a significant linear trend, $F\ (1, 16) = 60.70$, $p < .01$. The quadratic trend was not significant. The linear component accounted for 99.9% of the variance due to the number of rules.

Discussion

The striking differences in latencies produced by the different number of rules being applied is further support for the conclusion that rules are being used by *S*s in this task. The linear function of the increase in latencies indicates that each rule takes a constant amount of time, regardless of whether it is the first, second, or third rule to be applied to a string. However, since *S*s were instructed to apply the rules serially, this result does not exclude the possibility that these rules could be applied in other manners which would more closely approximate a model of parallel rather than serial processing.

At least one feature of behavior in the present experiment was obviously vastly different from language behavior and thus requires a different psychological model. The latencies demonstrate that in the experimental task it takes nearly 5 sec. to perform one permutation transformation. Since the analysis of even the simplest sentences in natural language involves many transformations, clearly the latencies involved in performing natural language transformations cannot be of even the same order of magnitude as they are in the experimental situation.

REFERENCES

Anderson, N. S. Word associations to individual letters. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 541–545.

Bourne, C. P., & Ford, D. F. A study of the statistics of letters in English words. *Information and Control,* 1961, **4**, 48–67.

Chomsky, N. *Syntactic structures.* The Hague: Mouton, 1957.

Chomsky, N. *Aspects of the theory of syntax.* Cambridge, Mass.: M. I. T. Press, 1965.

Jenkins, J. J., & Palermo, D. S. Mediation processes and the acquisition of linguistic structure. In U. Bellugi & R. Brown (Eds.), The acquisition of language. *Monographs of the Society for Research in Child Development,* 1964, **29** (2, Whole Number 92).

Martin, J. G., & Strange, W. Determinants of hesitations in spontaneous speech. *Journal of Experimental Psychology,* 1968, **76**, 474–479.

Miller, G. A. Project grammarama. In, *The psychology of communication: Seven essays.* New York: Basic Books, 1967.

Mowrer, O. H. *Learning theory and the symbolic process.* New York: Wiley, 1960.

Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford University Press, 1953.

Shaw, R., & Pufall, P. A push-down store model for patterned sequence learning. Paper presented at the meeting of the Mathematical Psychology Association, Stanford, August 1968.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received January 11, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 34–37

SELECTIVE ATTENTION IN MULTISOURCE MONITORING TASKS[1]

PETER HAMILTON[2]

University of St. Andrews

The monitoring of three information sources was studied using the method of observing responses. In the first stage of the experiment, four groups of four *S*s were assigned to a 2 × 2 block having two conditions of observing rate (paced and unpaced) and two fault rates (fast and slow). The probabilities of faults on the three sources were .6, .1, and .3. At the slow fault rate, unpaced *S*'s observing of the sources was independent of source probability, but became structured more in favor of source probabilities when fault rate was increased or when observing rate was slowed by pacing. The design was expanded by adding one faster fault rate condition and two slower pacing conditions to give a 4 × 3 block with four *S*s per cell. The trend towards selection on the basis of probability reached an asymptote where sources were observed with probabilities about .46, .22, and .32. It appeared that the variable governing the change in observing was the probability of a fault on any observation, $P(\mathrm{F/O})$, which is affected by both fault rate and pacing changes. A function relating $P(\mathrm{F/0})$ to the change in observing for pacing changes was derived, and the corresponding data for fault rate changes were found to fit that function.

In multisource monitoring situations, *O* must continuously select information from different sources. A common example of such a task is the maintenance of a number of dial readings within set limits. The processes underlying the deployment of attention over various sources are not yet fully understood. Senders (1965) showed that if the information load of the task was near *O*'s maximum capacity, the frequency of fixation of each source was a function of its bandwidth (signal probability). Hickey and Blair (1959) and Blair and Kaufman (1959) found no relationship; frequency of fixation of a source was independent of signal probability.

The present experiment was concerned with monitoring (sampling) behavior in a simple three-source task. It was designed so that the locus of attention at any moment was known. The technique used was developed by Holland (1958) and requires *S* to make a motor response to obtain a brief glimpse of the source he wishes to sample. Any one of three signal sources may be exposed by pressing one of three corresponding observation buttons. An assessment of the relative attractiveness of each source is obtained by recording *S*'s observing responses.

Pilot tests showed that changes in observing behavior might be produced by variation of signal rate or by pacing *S*'s observing responses. These variables were manipulated in the present experiment to assess their importance as determinants of monitoring performance. The experimental situation was an analogue of a real-life fault detection task. The *S* was required to monitor three information sources that could assume two states, "normal" and "fault." Faults remained in the present until seen and corrected, and *S*'s task was to find and correct the faults as soon as possible after they occurred. The experiment was conducted in two stages.

GENERAL METHOD

Subjects.—The *S*s were 48 undergraduates who were paid for their participation. Of these, 16 took part in Stage 1 and 32 in Stage 2.

Apparatus.—The *S* was seated in a darkened room at a control panel facing a horizontal bank of three lights, 1 ft. apart at a distance of 10 ft. Each light was a combination unit capable of flashing red, blue, or green, or any combination of these. At the control panel were three "sampling buttons," each corresponding to one of the lights, which when pressed caused the light to flash for 50 msec. The light under "normal" conditions was a red-and-green combination, and a "fault" condition was indicated by the appearance of a red-and-blue combination on illumination of the appropriate light. Also on the control panel were three corrector buttons. Pressing one of these would correct a fault on the corresponding light.

---

[1] This research is part of a thesis submitted in partial fulfillment of the requirements for the PhD degree at the University of St. Andrews, Scotland.

[2] Requests for reprints should be sent to Peter Hamilton, who is now at the Medical Research Council, Applied Psychology Research Unit, 15 Chaucer Road, Cambridge, England.

In an adjacent room, faults were programmed into the system by a Creed tape reader, which gave a momentary pulse for each fault. The tape reader could be cam-driven at speeds in the ratio 1:2:4. The momentary pulse was converted to a steady state by a holding circuit, so that any fault programmed into the system remained there until discovered and corrected by *S*. Provision for pacing *S*'s observing responses so as to control his sampling rate was made by installation of a bell and variable-speed revolving cam. Two four-channel Rustrak event recorders were used to record each sample made by *S* and the onset and correction of faults on each source.

Instructions.—The *S*s were asked to imagine themselves in the position of an operator on machine maintenance with three machines to look after. The state of each machine could be checked by flashing the relevant indicator light, and if this showed red and blue, a fault had occurred. The method of correction was then demonstrated. It was stressed that faults would occur and would remain in the system until discovered and corrected and that *S* must try to catch each fault as soon as possible after it had occurred. The *S*s were told that they could press any button they wished, but only one at a time.

STAGE 1

Method

Design.—A 2 × 2 design employing a slow and a fast fault rate condition and paced and unpaced observing conditions was used. To each of the four cells, four *S*s were allocated at random. In the unpaced condition, *S* was left to adopt his own observing rate; pacing was effected by ringing the bell in *S*'s room every 1.5 sec. and allowing one observing response only on each ring. In the slow fault rate condition, there was an average interfault interval of 30 sec. randomized between 16 and 44 sec. In the fast condition, there was an average interfault interval of 15 sec. randomized between 8 and 22 sec. Labeling the lights from left to right, A, B, and C, the probabilities of fault occurrence were: p (A) = .6, p (B) = .1, p (C) = .3. Within these probability restrictions, the series was constructed so that no sequential dependencies existed. No feedback was given.

TABLE 1

MEAN PERCENTAGE OF OBSERVING RESPONSES MADE TO EACH OF THE THREE LIGHTS IN EACH OF THE FAULT RATE AND PACING CONDITIONS

Fault rate	Pacing					
	Unpaced			Paced		
	A	B	C	A	B	C
Slow	27.6	40.6	31.8	40.8	25.9	33.3
Fast	38.6	30.1	31.3	45.5	23.3	31.2

Note.—For each cell, N = 4. A, B, and C = lights.

Procedure.—The *S*s were required to attend two 40-min. sessions a day apart. At the end of the second session, each *S* was asked to assess the probability of a fault on each source by guessing how many faults out of 100 would be on Sources A, B, and C.

Results

The *S*s in the slow fault rate condition processed 160 faults. To control any differential influence of learning, all *S*s' observing behavior at 160 faults was taken as the index of their performance, irrespective of fault rate. For each *S* in each group, 100 observing responses before and after the one hundred sixtieth fault were extracted from the record, giving a sample of 200 responses. The number of observations of Sources A, B, and C in the sample was expressed as a percentage (percentage of A, percentage of B, percentage of C).

The obtained observing frequencies for the three lights are shown in Table 1.

For statistical treatment of these data, an index is required of the degree to which *S*s discriminated between sources on the basis of the probability of a fault on these sources (selectivity). The index chosen was frequency of observing the high-probability source A (percentage of A). An overall analysis of variance on percentage of A indicated significant effects of pacing, F (1, 12) = 12.56, $p < .01$, and fault rate, F (1, 12) = 9.49, $p < .01$. The Fault Rate × Pacing interaction was not significant.

It appears from Table 1 that selectivity is increased as fault rate is increased and when pacing is introduced. The effect of

TABLE 2

MEAN PERCENTAGE OF OBSERVING RESPONSES TO EACH OF THE THREE LIGHTS FOR FOUR PACING AND THREE SPEED CONDITIONS

Fault rate[a]	Unpaced			Paced								
				1.5 sec.			2.5 sec.			4 sec.		
	A	B	C	A	B	C	A	B	C	A	B	C
1	27.6	40.6	31.8	40.8	25.9	33.3	47.3	23.5	29.2	46.4	22	31.6
2	38.6	30.1	31.3	45.5	23.3	31.2	45.5	21	33.5	45.8	20.6	33.6
3	43.4	24.6	32.1	44.9	25.9	29.2	42	27.7	30.3	45.6	24.8	29.6

Note.—For each cell, $N = 4$. A, B, and C = lights.

[a] Ratio of Fault Rates 1, 2, and 3 is 1:2:4.

pacing was to slow the rate of observing, the mean number of observations per minute for unpaced *S*s being 128.1 as compared to the imposed 40 per minute of the paced groups. It is not clear whether a limit to the trend of increased selectivity has been reached.

STAGE 2

Method

In Stage 2, the experiment was expanded in an attempt to locate the asymptote of the process.

Design.—The 2 × 2 block was expanded by adding two conditions of increased pacing restriction and one increased fault rate condition to produce a 4 × 3 block. The added pacing rates were 2.5 and 4 sec., and the third fault rate produced an average interfault interval of 7½ sec. randomized between 4 and 11 sec. All other conditions of the experiment remained unchanged. Four *S*s were allotted randomly to each of the eight additional blocks.

Results

The obtained observing frequencies for the three lights in all 12 cells are shown in Table 2. Again the precentage of observations of the high-probability light was used as an index of selectivity. Analysis of variance now indicated a significant interaction, $F\ (6,\ 36) = 2.61,\ p < .05$, and the pacing and fault rate effects were not significant. The indication is that the effect of an increase in fault rate depends on the pacing rate and vice versa.

The results of *S*s' assessments of fault probability are shown in Table 3 and indicate that probability learning was stable and accurate after 160 faults in all the slow fault rate conditions.

DISCUSSION

When fault rate is increased or the pacing interval lengthened, the number of observations possible per interfault interval is decreased. Thus, it may be that probability of a fault on any observation, $P(\text{F/O})$, is the main variable in the present experiment and increase in fault rate or pacing interval are effectively the same thing. To examine this proposition, the data in Table 2 may be collapsed onto the speed axis and the pacing axis. It is then possible to derive a measure of percentage-of-A samples (the selectivity measure) as a function of number of observations per interfault interval for both rate and pacing changes. Figure 1 shows that rate and pacing changes give rise to the same function. Since Senders (1965) and Blair and Kaufman (1959) used different levels of $P(\text{F/O})$, the latter having the lower level, it is likely that the two sets of data represent different points on the selectivity continuum shown in Fig. 1. It is interesting that the relationship between $P(\text{F/O})$ and selectivity found here may be found in the form of a rise in detection rate with increasing signal event/nonsignal event ratios in vigilance (Colquhoun, 1961).

TABLE 3

MEAN ASSESSED PROBABILITY OF FAULTS ON THREE LIGHTS FOR ALL FAULT RATE CONDITIONS

Fault rate	Mean assessed probability		
	A	B	C
1	.59	.09	.32
2	.59	.11	.30
3	.56	.12	.32

Note.—Assessments made after 160 faults at Rate 1, 320 faults at Rate 2, and 640 faults at Rate 3. A, B, and C = lights.

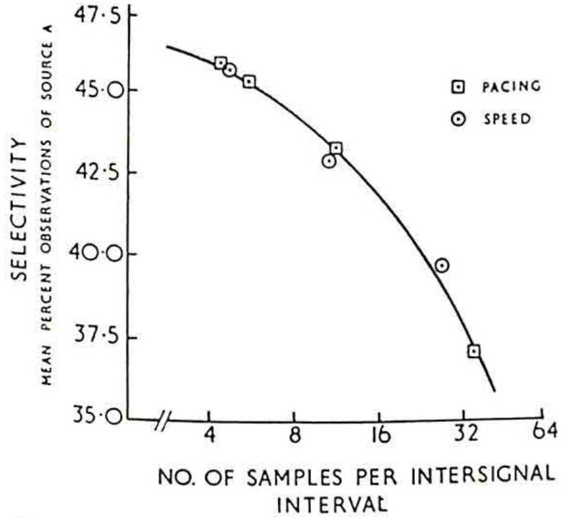


FIG. 1. Percentage of observations of Source A as a function of number of samples per intersignal interval.

Whether *O*s in a real-life monitoring situation will behave as the experimental *S*s is not yet known. From these data, it would appear that *S* chooses a high physical work rate in the unpaced situation and in some way releases himself from the burden of structuring his observing on the basis of probability, at least at low signal rates. Two unpaced *S*s observed much more slowly than the rest and exhibited higher selectivity. The analysis of monitoring behavior in terms of information theory (Senders, 1965) has ignored such strategy factors and concentrated on those situations where behavior is constrained by the high-information load of the task. In the light of these present data a reappraisal of current theory is required.

REFERENCES

BLAIR, W. C., & KAUFMAN, H. M. Multiple display monitoring: II. Control-display spatial arrangement. (Tech. Rep. No. SPD59-082) Groton, Conn.: General Dynamics Corporation Electric Boat Division, 1959.

COLQUHOUN, W. P. The effect of 'unwanted' signals on performance in a vigilance task. *Ergonomics,* 1961, 4, 41–51.

HICKEY, A. E., & BLAIR, W. C. 1. Multiple display monitoring. (Tech. Rep. No. SPD59-002) Groton, Conn.: General Dynamics Corporation Electric Boat Division, 1959.

HOLLAND, S. G. Human vigilance. *Science,* 1958, 128, 61–68.

SENDERS, J. W. An investigation of the visual sampling behaviour of human observers. (Tech. Rep. No. 1246, Contract No. NASI-3860) Cambridge, Mass.: Bolt Beranek & Newman, 1965.

(Received January 13, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 38–45

DISCRIMINATION, CONTRAST, AND CHAINING EFFECTS OF PRIOR TRAINING WITHOUT DISCRIMINANDA AND RESPONSE-CONTINGENT DELAY OF DISCRIMINANDUM PRESENTATION[1]

JOHN R. PLATT,[2] PETER C. SENKOWSKI, AND ROBERT MANN
University of Iowa

Two groups of rats were administered differential conditioning in a discrete-trials lever-pressing situation following either minimal or extended amounts of consistently rewarded pretraining in the absence of either discriminandum. On random trials, presentation of the discriminandum was delayed until various numbers of the required responses had occurred. It was found that extensive pretraining resulted in more rapid development of differential responding and higher asymptotic response speeds to both discriminanda. Positive behavior contrast was observed relative both to conventional, predifferential base lines and to those provided by responding during absences of discriminanda in differential conditioning. Little evidence was found for losses in discriminative stimulus control when the discriminandum presentation was delayed until after responding had begun. These results were discussed in the context of frustrative nonreward theory and the response chaining hypothesis.

Behavior theorists have long been interested in differential conditioning or single-presentation discrimination learning as a technique for simultaneously investigating acquisition and extinction processes, as well as their interactions. One result emerging from this research has been the finding of functional relationships between amount of predifferential training and rate of acquiring a discrimination. In a discrete-trials situation, using a differential conditioning technique, it has been found that discrimination develops more rapidly as a function of increasing numbers of prior continuous rewards (Amsel & Ward, 1965, Exp. II). A study by Shoemaker (1953) indicates that this relationship also holds for selective learning or simultaneous presentation discrimination.

This relationship between predifferential training and development of a discrimination is of considerable systematic importance since it is a major prediction of frustrative nonreward theory (Amsel, 1962). According to this theory, increasing numbers of predifferential rewards, to either or both of the discriminanda, increase the fractional anticipatory reward response (r_R) initially evoked on nonrewarded trials so as to increase the amount of frustration (R_F) elicited by the nonreward event. Since the stimulus consequences (s_F) of anticipatory forms (r_F) of R_F are assumed to evoke competing responses (avoidance) which are at least in part responsible for decrements in performance to the nonrewarded discriminandum (S−), greater numbers of prior rewards would be expected to hasten discrimination.

Unfortunately, both of the studies which provide evidence for this prediction have confounded predifferential reward history with simple exposure to the discriminanda. This confounding is particularly critical since a rather large body of literature (Epstein, 1967, Ch. 5) suggests that nondifferential exposure to a stimulus enhances its subsequent function as a discriminandum, even if the prior exposure is not rewarded. One purpose of the present study was to see if the Amsel and Ward (1965) result would occur when predifferential training did not involve presentation of either of the stimuli which were to serve as discriminanda. If such a result were to be obtained under the conditions just described, simple prior exposure to the discriminanda would be eliminated as a necessary condition for the phenomenon, while the frustration interpretation would remain viable since conditioning of r_R to the nondifferential situational stimuli should be sufficient to produce the predicted effect. This finding would, of course, have no bearing on the role of simple exposure to the discriminanda as a sufficient condition for acceleration of discrimination learning.

---

[1] This research was supported in part by United States Public Health Service Grant MH-15380-01 (John R. Platt, principal investigator).

[2] Requests for reprints should be sent to John R. Platt, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

Yet another interpretative problem arises in respect to using Amsel and Ward's (1965) result to support frustration theory. On the basis of such an account, the faster discrimination in *S*s receiving more predifferential rewards should be the result of more rapid extinction in S−, over and above any enhancement of performance in the rewarded discriminandum (S+) as a result of the shift from continuous to intermittent reward at the beginning of differential conditioning. The well-known phenomenon of positive behavioral contrast (PBC), which has been reviewed by Dunham (1968), certainly suggests the possibility that Amsel and Ward's *S*s may have responded faster to S+ during differential conditioning than at the end of predifferential training. What is not clear is whether the magnitude of such PBC would be expected to vary as a function of number of predifferential rewards so as to account for some or all of the obtained differences in rate of discrimination.

Unfortunately, there appear to be no reported studies of the effects of amount of predifferential training on magnitude of PBC. This oversight probably stems from the fact that asymptotic or steady-state predifferential performance is usually used as the base line against which PBC is evaluated. Thus, manipulation of amount of predifferential training so as to prevent attainment of stable performance precludes assessment of PBC in the usual manner. Platt (1967) found that response speeds comparable to the predifferential base line could be obtained in differential training during brief periods prior to the presentation of either discriminandum, early in a trial. The present study was designed to make use of this finding to obtain PBC estimates in both the presence or absence of extended predifferential training. Such estimates should serve to indicate what role PBC plays in differing rates of discrimination learning resulting from pretraining manipulations.

A final goal of the present study was to investigate the interaction of exteroceptive discriminanda and response-produced chaining stimuli. Although behavior theorists have long been fond of the notion that a response may be sequentially chained by being conditioned to its own stimulus consequences, only the most indirect evidence for such a mechanism has been forthcoming. Donahoe, Schulte, and Moulton (1968) investigated the effects of adding an exteroceptive stimulus gradient to response chains in which within-chain delay of reward either was or was not correlated with response speed. On the basis of failures to modify the response speed gradients in the former type of chain, these investigators concluded that its stimulus control was entirely of the interoceptive, chained variety. The overgeneralization that may be prompted by these results is that once such a chain is initiated, it is immune to further exteroceptive control. By manipulating the point of presentation of a discriminandum relative to the initiation of a response chain, the contribution of chained stimulus control should be directly reflected as a relative loss of control by the discriminandum when it is presented within the chain. Such a manipulation was made in the present study so as to further determine the existence and extent of chaining in a situation involving a correlation between response speed and within-chain delay.

METHOD

Subjects.—Thirty-two male albino rats of the Sprague-Dawley strain were randomly assigned to two groups. All *S*s were approximately 130 days old at the beginning of the experiment. Three weeks before, and throughout the experiment, all *S*s were maintained on a 12-gm/day food ration, including rewards obtained in the experimental situation.

Apparatus.—Two $11\frac{1}{2} \times 9\frac{1}{2} \times 10\frac{1}{4}$ in. test chambers of identical construction were used. Centered on one wall, $2\frac{1}{2}$ in. above the grid floor, was a Hawley retractable lever which required a static force of 38 gm. to operate. A stainless-steel shield extended 1 in. from this wall, immediately below and 2½ in. up one side of the lever opening. A similar shield was placed 1 in. from the other side of the lever opening and formed one side of the food cup which was immediately adjacent to the lever opening. An additional plate ¾ in. high was placed across the front of the shield, and a final piece of stainless steel separated the lever from the food cup. The resulting configuration exposed only the top surface of the lever, thus preventing the lever from striking *S*, as well as eliminating many variations in response topology.

Two small white pilot lamps were located 5¼ in. above the grid, 3 in. on either side of the lever. These lamps served to dimly illuminate the chamber, as well as to signal trials. Behind the chamber wall was located a speaker and a motor-driven pellet dispenser. The test chambers were placed in sound attenuating enclosures and were successively controlled by solid-state circuitry located in an adjacent room. Response times and experimental parameters were recorded on punched paper tape with a resolution and minimum value of .1 sec. (Platt & Senkowski, 1968).

Procedure.—The first three sessions of pretraining were identical for all *S*s. In the first session, each *S* was given 30 magazine training trials. On each trial, three 45-mg. Noyes pellets were presented at the end of a 2.5-sec. warning signal. The warning signal consisted of an increase in illumination from approximately .1 ftc. to approximately 2.2 ftc. The interval between reward presentations was 55 sec. In the second session, 5 magazine training trials were followed immediately by 25 lever-pressing trials. A lever trial began with presentation of the warning signal, followed 2.5 sec. later by entry of the lever into the chamber. When *S* pressed and released the lever, a three-pellet reward was accompanied by retraction of the lever and termination of the warning signal. If no response occurred within 25 sec. of lever presentation, the trial was terminated without reward. Trials were initiated at the rate of 1 every 55 sec. The third session was similar to the second, except the number of responses required per trial was gradually increased to seven.

Sixteen *S*s comprising Group S (separate phase) were given 40 additional training sessions. These sessions were the same as described previously, except that seven responses were required for reward on all 30 trials. Following this additional predifferential training, 44 sessions of differential conditioning were administered. Throughout differential conditioning, half of the trials were nonrewarded according to predetermined irregular orders. These two kinds of trials were differentiated by 4,000- and 10,000-cps tones, counterbalanced between *S*s. The 4,000-cps tone was presented at 80 db. (SPL), and the 10,000-cps tone at 70 db. The background noise level was 68 db. On 5 of each of the 15 reinforced trials and 15 nonreinforced trials in each session, the tone was presented at the onset of the warning signal (WS). On another 5, it was presented after the first lever press (R_1), and on the remaining 5, after the fourth press (R_4). In each case, the tone continued until the trial was terminated, and the various delays of tone presentation occurred in a predetermined irregular order.

The remaining 16 *S*s, comprising Group M (mixed phase), began differential conditioning immediately after the third pretraining session and also continued for 44 sessions. For each lever press throughout the experiment, an interresponse time (IRT) was recorded as the interval between successive releasings of the lever or, in the case of the first press of a trial, as the time between entry of the lever into the chamber and *S*'s first releasing of it.

RESULTS

Mean gradients of response speed over the seven lever presses in a trial are shown in Fig. 1 for early stages of differential conditioning in both Groups S and M. Later stages are portrayed in Fig. 2. Gradients for rewarded (S+) and nonrewarded (S−) trials under each of the three conditions of discriminandum delay were plotted separately. In order to evaluate the statistical reliability of effects apparent in this data, an analysis of variance was performed in which the main effects were amount of pretraining, successive stages of differential conditioning, discriminanda, intratrial delays of discriminandum presentation, and ordinal response position in the trial gradient. A .01 significance level was adopted in view of the considerable power of the design.

The obvious development of reliable speed differences between S+ and S− was statistically supported by significant effects of discriminanda, *F* (1, 30) = 131.92, and Discriminanda × Stages, *F* (5, 150) = 71.76. It can be seen that Group S displayed considerable stimulus control by the second block of differential conditioning sessions, while the process was somewhat slower for Group M so as to produce a significant Discriminanda × Pretraining interaction, *F* (1, 30) = 10.85, and a Discriminanda × Pretraining × Stages interaction, *F* (5, 150) = 3.59. As a matter of fact, the differences between S+ and S− speeds never became as large in Group M as in Group S. However, Group M did eventually catch up to and slightly surpass Group S if discrimination was indexed by the ratio, rather than the difference, of S+ and S− speeds.

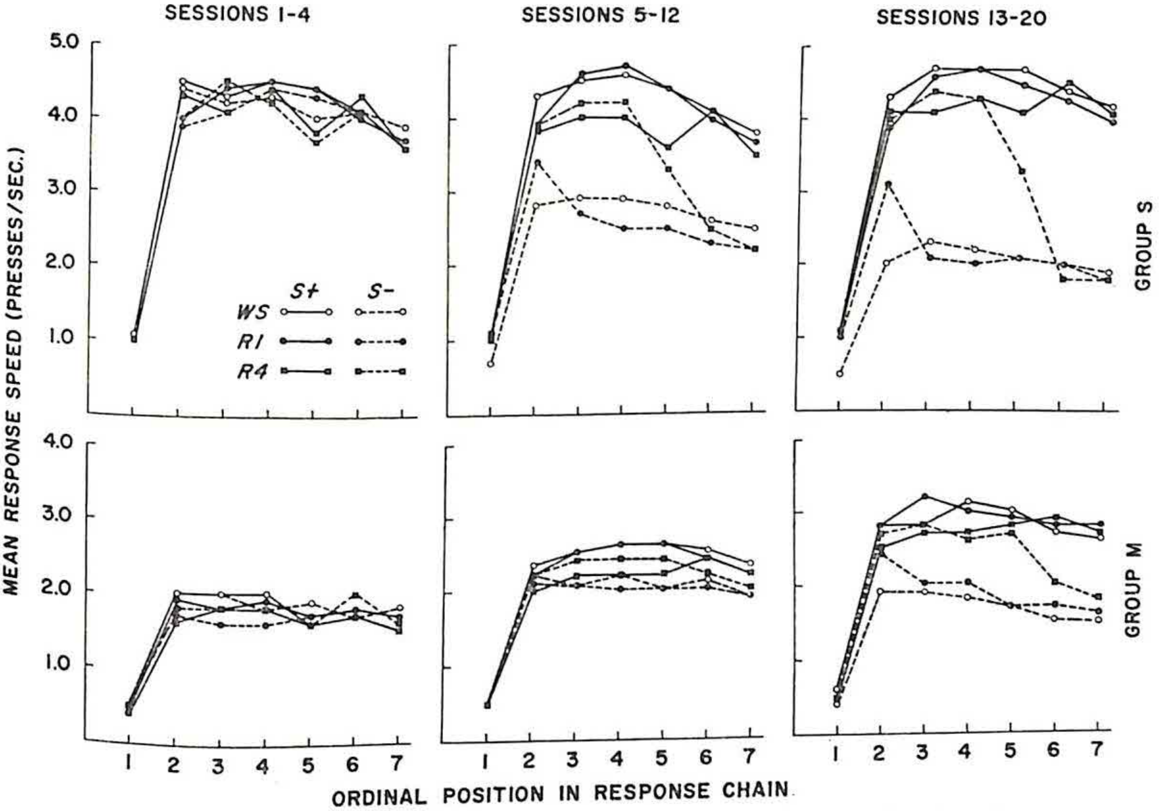


FIG. 1. Early differential conditioning response speed gradients as a function of discriminandum delay and amount of predifferential training.

Further insight into differences in the development of stimulus control in the two conditions of pretraining may be had by examining Fig. 3, which graphically represents the previously mentioned Discriminanda × Pretraining × Stages interaction. This figure makes clear that a large part of the more rapid emergence of discrimination in Group S resulted from the fact that stimulus control in this condition was represented almost entirely by decreases in S− speed. In Group M, on the other hand, S− speeds increased for a time before beginning to drop away from the still increasing S+ speeds.

The representation in Fig. 3 also serves to emphasize the overall faster response speeds of Group S as compared to Group M. This difference was reflected statistically by a significant effect of pretraining, F (1, 30) = 11.96, and by the Pretraining × Stages interaction, F (5, 150) = 10.14. Such a difference would have been expected early in differential conditioning on the basis of differing amounts of pretraining. However, the fact that this difference persisted at least 24 sessions after the S+ speeds of Group M had become asymptotic suggests a permanent effect of pretraining on overall response speed in differential conditioning.

A final aspect of the effects of pretraining on the development of discrimination occurred in respect to the degree of control exerted by the discriminanda as a function of their delays. Figures 1 and 2 clearly indicate that substantial control came to be exerted by the discriminanda at all delays of presentation. In view of the very rapid discriminandum-independent increase in response speeds from the first to the second response of a trial, it is difficult to compare degree of discriminative control at various presentation delays simply in terms of the rate of divergence of S+ and S− gradients. Nevertheless, a significant Delays × Gradient × Pretraining interaction, F (12, 360) = 5.78, and a Stimuli × Delays × Gradient × Pretraining interaction, F (12, 360) = 8.14, statistically support the observation

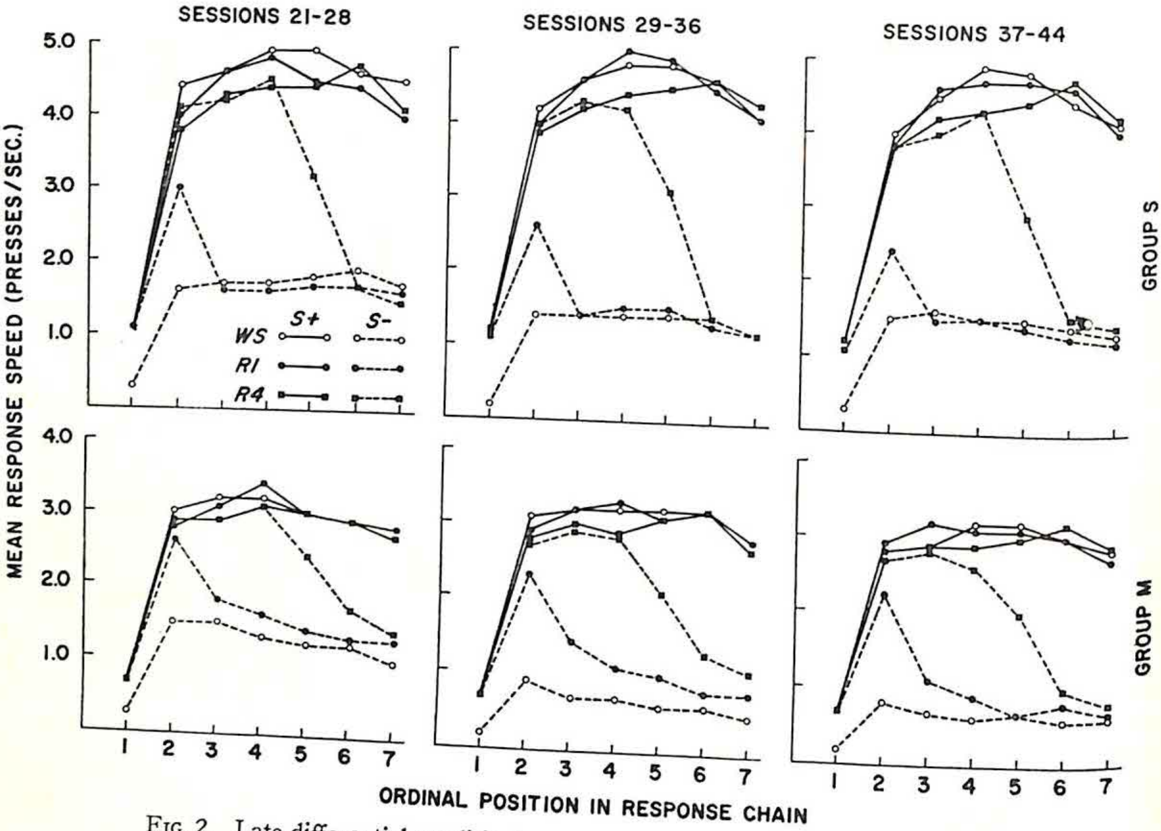


FIG. 2. Late differential conditioning response speed gradients as a function of discriminandum delay and amount of predifferential training.

from Fig. 1 and 2 that Group S showed more rapid divergence of S+ and S− gradients with shorter delays of discriminandum presentation, while slightly the reverse was true for Group M. Evaluation of these effects is further complicated by the fact that early in differential conditioning, Group S appeared to actually approach lower terminal levels of response speed with longer delays of S− presentation.

Several other significant sources of variance emerged from the overall analysis of variance, but were either of no systematic interest or were of such complexity as to elude interpretation.[3] Further results are more profitably considered in respect to PBC. The present experiment offered two possible base lines for evaluating PBC. Of these two, only the speed of responses prior to discriminandum presentation during differential conditioning was available for both Groups S and M. In order to facilitate evaluation of possible contrasts against this base line, the analysis of variance described earlier was repeated using only the data of S+ trials. This seeming redundancy served both to greatly simplify interpretation of complex sources of variance and also to increase homogeneity of variance. A significance level of .01 was again adopted.

Referring to Fig. 1 and 2, the observation of faster responding after than before S+ presentation was statistically reliable, as indicated by significant effects of delay, F (2, 60) = 20.22, and Delays × Gradient, F (12, 360) = 11.14. Some statistical support was also forthcoming for the greater magnitude of this effect in Group S since the Delays × Gradient × Pretraining interaction, F (12, 360) = 2.45, was significant, although Delays × Pretraining was not. Figure 1 also suggests that the difference between response speeds before and after S+ presentation did not appear until Ss displayed some discrimination between S+ and S−, although stages did not interact significantly with delays or with Delays × Groups. Other significant effects indicated by the analysis have already been dealt with in the earlier presentation of discrimination results, except for the Pretraining × Gradient interaction, F (6, 180) = 5.66. This interaction reflects the greater flatness of the response speed gradients in Group M than in Group S.

---

[3] Complete summary tables of the analyses of variance reported in this paper are available from the authors.

The justifiability of using response speed in the absence of either discriminandum as a base line for evaluating PBC is considered in the Discussion section. The predifferential performance of Group S, however, does provide for an empirical evaluation of the similarity of this rather unusual base line to the customary one. Figure 4 shows the response speed gradient for Group S in the last session of predifferential training, as well as this group's gradient for each of the three delays of S+ presentation over the last 24 sessions of differential conditioning. Testing a statistical null hypothesis of no performance difference between the end of predifferential training and in the absence of discriminanda during differential conditioning would serve little purpose, but Fig. 4 clearly indicates that much the same conclusions concerning PBC would have been reached regardless of which had been used as a base line. Figure 4 also shows two additional results. First, the magnitude of PBC relative to the predifferential base line showed a decided increase as the end of a trial was approached. That this increase was not due to time since S+ presentation is indicated by the fact that by the second response after S+, response speeds had always attained the level generated by earlier presentations. Second, regardless of whether S+ had been presented, speeds of the first response in the behavior chain were faster in differential conditioning than at the end of predifferential training, F (1, 45) = 96.43. This last finding provides the single exception in the present study to the generalization that all PBC was totally under control of S+.

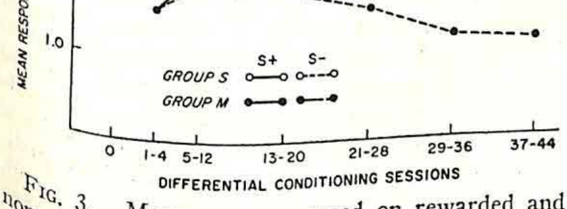


FIG. 3. Mean response speed on rewarded and nonrewarded trials as a function of stage of differential conditioning and amount of predifferential training.

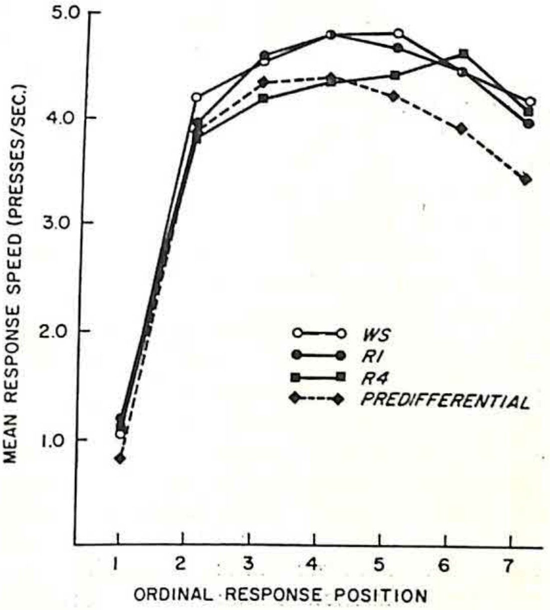


FIG. 4. Mean response speed gradients at the end of pretraining and for the three delays of S+ presentation over the last 24 sessions of differential conditioning in *S*s given extended pretraining.

DISCUSSION

Results of the present study clearly indicated that faster discrimination learning following greater amounts of consistently rewarded pretraining could not be attributed solely to prior experience with the discriminanda since neither discriminandum was presented during pretraining. In addition, the facilitative effects of pretraining on discrimination were found to be most apparent in the increased rate at which speeds in S− decreased, rather than to any enhancement of PBC in S+. This latter fact can be most easily seen by referring back to Fig. 1 and comparing performance in S+ and in S− to the prediscriminandum portion of the trials for both of the groups during the early stages of discrimination.

Both of the results just described are in agreement with Amsel's (1962) frustrative nonreward account of the effects of predifferential training on discrimination, as discussed previously in the present paper. In terms of this theory, the present results are also in agreement with the finding of more rapid discrimination of differential reward magnitudes as a function of larger rewards during predifferential training (Platt & Gay, 1968). This parallel results from the assumption that number and magnitude of rewards function similarly to increase r_g, while nonreward and small reward function similarly to elicit R_F in the presence of large r_gs.

The major exceptions to these generalizations appear to occur in free-operant studies. Thus Terrace (1963) found that 21 prior sessions of variable-interval (VI) reward in S+ greatly increased the number and rate of responses in a subsequently introduced S− which was never rewarded. Taus and Hearst (1968) recently reported that under Terrace's conditions, rate of discrimination is actually a nonmonotonic function of amount of prior VI training in S+, so that the fewest S− responses occur after an intermediate amount of predifferential training.

An extremely large number of differences exist between the free-operant studies just described and discrete-trials experiments such as the present one. Nevertheless, certain points of discrepancy may be singled out as promising factors in attempting to account for the different results which have been obtained under these two types of procedures. In the first place, discrete-trials procedures typically involve a number of distinctive interoceptive and exteroceptive cues which accompany the termination of a trial and the attendant reward. Within this context, consistent reward training means that a perfect positive correlation is established between these terminal stimuli and reward. Since such exteroceptive cues are, in general, minimized by a free-operant VI procedure and any remaining interoceptive cues are randomized with respect to reward, it is quite possible that free-operant VI training is much more like intermittent than consistent reward in the discrete-trials case. If this analogy is correct, then results such as Terrace's (1963) could be easily understood as a partial reinforcement extinction effect generalized from the predifferential stimuli to those present on S− trials. Such effects have been treated in terms of frustrative nonreward theory (Amsel & Ward, 1965).

Another typical difference between free-operant and discrete-trials paradigms is related to the fact that any attempt to study the degree of emotional or aversive properties of nonreward in discrimination learning by manipulating predifferential training may also produce differences in the probability of S− responding at the onset of discrimination training. This problem would seem to be particularly acute in free-operant situations with time-contingent discriminandum exposure. Studies such as Terrace's (1963) involved great differences in the density of nonreward reponding at the beginning of discrimination training. On the other hand, discrete-trials paradigms typically use response-contingent discriminandum exposure which serves to hold the number of nonreward responses constant until *S*s refuse to respond to S− and a time criterion has to be imposed. Response-contingent discriminandum exposure should probably be preferred if one is attempting to test theoretical propositions concerning the severity of nonreward, rather than its frequency of occurrence.

Turning now to the behavioral contrast phenomena found in the present study, it has already been indicated that such effects did not suffice to account for the obtained differences in discrimination. In addition, PBC alone was not sufficient to account for the overall faster response speeds in Group S at the end of discrimination learning since the speeds of Group M never approximated the predifferential asymptote for Group S.

In spite of the seemingly inconsequential role of PBC in development of the current discrimination the present results, along with those of Platt (1967), appear to provide the first clear demonstration of PBC in a discrete-trials discrimination situation. Although the Amsel (1962) double-runway frustration effect may be so considered to the extent that it is viewed as a discrimination between the first and second runways (Daly, 1968).

The major objective of the present study in respect to PBC was to investigate the usefulness of the delayed-discriminandum technique for establishing PBC base lines. The technique was largely supported in that prediscriminandum performance was comparable to predifferential performance in Group S, except for the speed of initiation of the response chain. Some caution should, of course, be exercised in using this result to justify the use of prediscriminandum performance as a PBC base line in conditions such as Group M which in principle do not allow corroboration of its equivalence to a prediscrimination base line. Further justification for this extension may be had, however, if it is pointed out that the prediscriminandum stimuli in the present study received 50% reward. Since in discrete-trials situations, 50%

reward has been found to produce response speeds comparable to or slightly greater than those produced by 100% reward (Goodrich, 1959), the prediscriminandum base line would at least not be expected to overestimate PBC.

It should be indicated that PBC observed in the present study may be treated within the same general theoretical framework as were the pretraining effects. Various investigators (Amsel & Roussel, 1952; Terrace, 1963) have explained increases in response speeds in one stimulus, following reductions in reinforcement frequency in another stimulus, in terms of motivational factors such as R_F. In the current results, both the sustained nature of PBC and its increasing magnitude toward the end of a behavior chain suggest that motivational properties of r_F generalized from S− would have to play an even more prominent explanatory role than R_F carrying over from a preceding nonreward trial. This explanation contains the apparent contradiction of crediting r_F-s_F with incremental effects on response speeds in S+, but decremental ones in S−. This problem may be resolved if it is remembered that there is ample evidence to indicate that the avoidance reactions initially evoked by s_F are counterconditioned by approach responses in S+ (Amsel & Ward, 1965).

The results obtained in the present study were not entirely unambiguous in respect to potential decrements in discriminative control resulting from delayed presentations of the discriminanda. Clearly, the response chain was not immune to control by exteroceptive discriminanda presented after its initiation since substantial discrimination was obtained for all delays of discriminandum presentation. The finding that any decrements in control that did result from delaying discrimination presentation occurred only in Group S might have been expected since the extended predifferential training given this group should have led to more complete chaining than in Group M. The rather minimal evidence for any sort of chaining in the present results suggests a reevaluation of the original rationale for using loss of discriminative control as an index of chaining. It might be that under conditions of variable discriminandum delay, *S*s made a series of chained or nonchained responses until the discriminandum was obtained and only then initiated a final chain if S+ occurred. Such a process would, of course, be even more likely in Group M than in Group S since the former should not have entered differential conditioning with well-established chaining. If this interpretation proves to be correct, the usefulness of loss of discriminative control as an index of chaining may be limited to situations in which the discriminanda are only infrequently delayed.

REFERENCES

Amsel, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, 1962, **69**, 306–328.

Amsel, A., & Roussel, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *Journal of Experimental Psychology*, 1952, **43**, 363–368.

Amsel, A., & Ward, J. S. Frustration and persistence: Resistance to discrimination following prior experience with the discriminanda. *Psychological Monographs*, 1965, **79**(4, Whole No. 597).

Daly, H. B. Excitatory and inhibitory effects of complete and incomplete reward reduction in the double runway. *Journal of Experimental Psychology*, 1968, **76**, 430–438.

Donahoe, J. W., Schulte, V. G., & Moulton, A. E. Stimulus control of approach behavior. *Journal of Experimental Psychology*, 1968, **78**, 21–30.

Dunham, P. J. Contrasted conditions of reinforcement: A selective critique. *Psychological Bulletin*, 1968, **69**, 295–315.

Epstein, W. *Varieties of perceptual learning.* New York: McGraw-Hill, 1967.

Goodrich, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *Journal of Experimental Psychology*, 1959, **57**, 57–64.

Platt, J. R. Intratrial effects in separate-phase differential instrumental reward conditioning. Unpublished doctoral dissertation, University of Texas, 1967.

Platt, J. R., & Gay, R. A. Differential magnitude of reward conditioning as a function of predifferential reward magnitude. *Journal of Experimental Psychology*, 1968, **77**, 393–396.

Platt, J. R., & Senkowski, P. C. A computer-compatible technique for recording binary events in real-time. *Behavioral Research Methods and Instrumentation*, 1968, **1**, 79–81.

Shoemaker, H. A. The relative efficiency of rewarded and nonrewarded training in a black-white discrimination problem. Unpublished doctoral dissertation, University of Colorado, 1953.

Taus, S. E., & Hearst, E. Operant discrimination learning after different amounts of nondifferential S+ training. Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1968.

Terrace, H. S. Discrimination learning with and without errors. *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 1–27.

(Received January 13, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 46–51

EFFECTS OF OVERTRAINING ON THE ACQUISITION OF INTRADIMENSIONAL AND EXTRA-DIMENSIONAL SHIFTS[1]

BRYAN E. SHEPP[2] AND FRANK D. TURRISI[3]

Brown University

Three groups of *S*s were given 0%, 100%, or 300% overtraining prior to transfer to a second discrimination. In the second discrimination, half of the *S*s in each overtraining group were trained on an intradimensional (ID) shift and half were trained on an extradimensional (ED) shift. ID shifts were superior to ED shifts, and an Amount of Overtraining × Shift interaction was observed. With an increasing amount of overtraining, ID shifts became progressively easier, whereas ED shifts became progressively more difficult. The results were interpreted as supporting an assumption of two-stage theories of discriminative learning, which implies that the probability of attending to the relevant stimulus dimension increases with increasing amounts of overtraining.

Several chaining or two-stage theories of discriminative learning have assumed that *S*s, in learning a discrimination, learn to attend to the relevant stimulus dimension and learn to approach the correct cue on that dimension (Goodwin & Lawrence, 1955; Lovejoy, 1966; Sutherland, 1959; Zeaman & House, 1963). These theories assume that both attention and instrumental responses are transferred to subsequent discriminations, and several varieties of transfer situations have been used to establish the validity of such views.

The most direct test of two-stage theories is the experimental comparison of an intradimensional (ID) with an extradimensional (ED) shift. Prior to transfer, *S*s are trained on a discrimination in which, e.g., form (e.g., $\overset{+}{\text{square}}$ and $\overset{-}{\text{circle}}$) is relevant and color (e.g., $\overset{\mp}{\text{red}}$ and $\overset{\pm}{\text{green}}$) is irrelevant. Following the acquisition of this problem, *S*s are trained on a second discrimination in which new forms (e.g., triangle and cross) and new colors (e.g., blue and yellow) are presented. Half of the *S*s are trained on an ID shift in which form and color are still relevant and irrelevant, respectively. The remaining *S*s are trained on an ED shift in which form is irrelevant and color is relevant. During acquisition of the initial discrimination, all *S*s are assumed to have learned to attend to the form dimension. Since form is relevant to the solution of the ID shift, positive transfer of the previously learned attentional response is arranged, and *S*s are required only to learn which of the two new forms is correct. However, since color becomes relevant in the ED shift, transfer of the attentional response is negative. To solve the ED shift, *S*s must extinguish the previously learned attentional response and learn the appropriate attentional and instrumental responses. Given these transfer arrangements, chaining theories predict that an ID shift will be learned faster than an ED shift. The results of several studies have confirmed this prediction (Campione, Hyman, & Zeaman, 1965; Furth & Youniss, 1964; House & Zeaman, 1962; Shepp & Eimas, 1964.)

Two mathematical versions of these two-stage theories predict details of learning and transfer data which other theories do not predict. One such version is the "one-look" model proposed by Zeaman and House (1963), and the other is Lovejoy's (1966) "two-stage" model. For present purposes, only assumptions that are common to both models are considered. The models assume that a stimulus situation consists of a set of n relevant and irrelevant dimensions, each of which can elicit an attentional response (O_i) with a certain probability (Po_i). A further assumption is that only one of the competing responses, e.g., O_1, is effective on any given trial. Attention to a dimension exposes the specific cues (e.g., S_1, S_1') of that dimension, and an approach response to S_1 or S'_1 may be made with the conditional probabilities $Pr_{(1)}$ and $1 - Pr_{(1)}$, respectively.

---

[1] This research was supported by Research Grant HD 01349 from the National Institute of Child Health and Human Development. The authors wish to express appreciation for the cooperation and assistance of John Smith, Superintendent of Ladd School.

[2] Requests for reprints should be sent to Bryan E. Shepp, Department of Psychology, Brown University, Providence, Rhode Island 02912.

[3] Now at Florida State University.

An important feature of these models is the manner in which reinforcement and nonreinforcement are assumed to affect the probabilities of the attentional and instrumental responses. The probabilities of all attention responses in any given situation are assumed to change on every trial. The attentional response that is elicited on a trial is directly reinforced or extinguished, and since the probabilities of the attentional responses must sum to 1.0, the nonelicited attentional responses are indirectly reinforced or extinguished. If, e.g., O_1 is elicited, then $Po_{(1)}$ is increased by reinforcement or decreased by extinction. An increase in Po_1 leads to a decrease in the probabilities of attending to other dimensions. Analogously, a decrease in Po_1 results in an increase of attention to other dimensions. Unlike attention probabilities, not all instrumental response probabilities change on every trial. Response tendencies to specific cues are modified only on trials when the attentional response that exposes the cues is directly reinforced or extinguished. For example, the conditional probabilities of approaching S_1 or S'_1 are changed when O_1 is elicited, but are unaffected on trials when O_2 is elicited.

As a consequence of these assumptions, the models predict that the probability of the relevant attentional response approaches an asymptote more slowly than does the probability of approaching the positive stimulus. Thus with a weak criterion or a few overtraining trials, the probability of attending to the relevant dimension may be less than asymptotic value, although *S*s consistently choose the correct stimulus. However, with increasing amounts of overtraining, the probability of attending to the relevant dimension increasingly approximates an asymptotic value.

The experimental consequences of these assumptions are clear. As the amount of overtraining is increased, acquisition of an ID shift should become increasingly facilitated, whereas the acquisition of an ED shift should become increasingly more difficult. The present study tested these predictions of the two-stage models.

METHOD

Subjects.—The *S*s were 64 experimentally naive retardates who were institutionalized at the Joseph H. Ladd School. The *S*s were assigned randomly to experimental groups without regard to diagnostic category or sex. The mean IQ of *S*s was 39.54 (range: 27–56) and the mean mental age (MA) was 73.52 mo. (range: 47–109 mo).

Apparatus.—The *S*s were trained on a modified Wisconsin General Test Apparatus. The *E* and *S* sat facing each other at a table, with a one-way vision screen interposed. A stimulus tray was presented directly in front of *S* by pushing it through a space beneath the one-way screen. A stimulus object was placed over each of two foodwells separated by 12 in. from center to center. The entire apparatus, except the screen, was painted gray.

The stimuli were stereometric objects composed of all combinations of seven forms (triangle, square, circle, diamond, cross, T, and star) and seven colors (red, green, yellow, orange, blue, brown, and black). The objects were cut from $\frac{3}{4}$-in. pine and were 3 in. high and 3 in. wide at the point of maximum width. The stimuli were mounted vertically on gray bases.

Each correct response was reinforced with one piece of cereal, a "Crispy Critter."

Procedure.—Each *S* was taken to the laboratory and told that he was going to play the "cookie game." After *S* was seated in front of the apparatus, *E* baited one of the foodwells and, with both wells uncovered, pushed the tray in front of *S*. The *E* then said, "Find the cookie," pointing to the baited well if *S* did not retrieve the reward. When *S* could quickly retrieve the reward, the baited well was covered with a multidimensional (junk) object, and *S* was again instructed to find the cookie. The position of the reward was varied from trial to trial by a Gellermann series. When *S* could easily dislodge the object and retrieve the reward, discrimination training was begun.

All *S*s were trained on a two-choice visual simultaneous discrimination in which one dimension was relevant and one dimension was irrelevant. The irrelevant dimension was presented in a variable-within manner such that a dimension, e.g., color, was represented by two values, e.g., red and green, on each trial. During the training series, each irrelevant stimulus was presented half of the time on the right and half of the time on the left. In each position, each irrelevant stimulus was paired half of the time with the positive stimulus and half of the time with the negative stimulus. Half of the *S*s were trained with form as the relevant dimension and color as the irrelevant dimension; the remaining *S*s were trained with the converse arrangement. A pair of stimuli of each dimension was chosen randomly from the set to compose the discrimination for each *S*, and the correct stimulus of the problem was designated randomly.

On each training trial, *E* baited a food well, positioned an object over each well, and pushed the tray in front of *S*. The first response of each trial was verbally reinforced ("Good") or punished ("No") by *E*. A correction procedure was used, and when *S* obtained the reward, the tray was withdrawn and a new trial was begun. Each *S* was trained for 40 trials a day until a criterion of nine correct responses in any 10 trials was reached. The *S*s who did not learn in 200 trials were dropped from the experiment. The *S*s who learned the first discrimination were assigned randomly to one of three overtraining conditions. The criterion group (NOVT) received no overtraining and was trained on a transfer task on the day after criterion was reached. A second group (100-OVT) was given 100% overtraining, and the remaining group (300-OVT) was given 300% overtraining. The base used for the administration of overtraining was the number of trials required by a given *S* to reach criterion. For example, an *S* who required 50 trials to reach criterion was given 50 trials of overtraining as a member of 100-OVT or 150 trials of overtraining as a member of 300-OVT. On the day after each *S* finished overtraining, training on the transfer task was begun.

In the transfer task, the visual dimensions presented in the original discrimination were used again and the irrelevant dimension was presented again as a variable-within array. The stimuli of Problem 1 were replaced by new ones in the transfer task. Half of the *S*s in each overtraining group were trained on an ID shift and half were trained on an ED shift. In the ID shift the relevant and irrelevant dimensions of original learning are unchanged in the transfer problem. In the ED shift the relevant dimension during acquisition is made irrelevant and the previously irrelevant dimension is made relevant. Otherwise, the procedures used in acquisition were used in transfer, and, again, *S*s were trained to a criterion of either nine correct responses in 10 trials or 200 trials.

RESULTS

Sixteen of 64 *S*s trained on the original discrimination failed to learn within 200 trials. For 9 of these *S*s, color was relevant, and for 7 *S*s, form was relevant. A *t*-test comparison of errors to criterion between all *S*s who were trained with color relevant and all *S*s who were trained with form relevant revealed no significant difference, $t\ (62) = .23$, $p > .05$.

The remaining 48 *S*s who learned Problem 1 were evenly divided among the six experimental groups; 4 *S*s in each group were trained with form relevant, and 4 *S*s were trained with color relevant. All data were analyzed in a 2 (shift) × 3 (overtraining) × 2 (dimensions) factorial design.

MA and IQ analyses.—An analysis of variance was run on both the MA and IQ scores of *S*s. These analyses indicate that there are no reliable differences between conditions with respect to either MA or IQ, $p > .05$.

Original learning.—Original learning was analyzed by using both the number of trials to last error and the number of errors required by each *S* to reach criterion. No main effect or interaction approached significance in either analysis, $p > .05$.

Transfer.—The measures used in the analysis of the transfer data were the number of trials to last error and the number of errors to criterion. The distributions of these measures are similar, and since both measures reveal the same effects, only the analysis of trials to last error is reported. The 2 × 3 × 2 analysis of variance of these data revealed a significant main effect due to shift, $F\ (1, 36) = 4.5$, $p < .05$, and a comparison of the means indicated that ID shift *S*s required fewer trials to last error than did the ED shift *S*s. A marginally significant main effect was observed for dimensions, $F\ (1, 36) = 3.4$, $p < .10$; form problems are easier than color problems regardless of type of shift. The most important theoretical effect is revealed by the significant Shift × Overtraining interaction, $F\ (2, 36) = 5.7$, $p < .01$. The means of the trial to last error in order of increasing amounts of overtraining are 46.1, 35.9, and 1.8 for the ID groups and 41.3, 31.5, and 120.5 for the ED groups. Orthogonal comparisons for simple effects indicate that the ED-300 group showed negative transfer in comparison with the ED-0 and ED-100 groups, $t\ (30) = 3.3$, $p < .01$, whereas the latter groups did not differ reliably, $p > .05$. Orthogonal comparisons of the ID groups did not reveal any significant differences. Since an inspection of the data reveals a rather obvious difference between the ID-300 groups and the other ID groups, further tests for simple effects were made with the Mann-Whitney test. The ID-300 group differs reliably from the ID-0 and ID-300 groups, $p < .05$. A difficulty in evaluating the ID shift differences is associated with the small variance of the scores in the ID-300 group. A comparison of this variance with the largest variance yields a maximum $F\ (6, 7) = 387.8$, $p < .001$. A logarithmic transformation successfully normalizes the data, maximum $F\ (6, 7) = 10.2$, $p > .05$, and a $2 \times 3 \times 2$ analysis of variance of the transformed data shows that main effects due to dimensions and shift and the Shift × Overtraining interaction are significant. More importantly, orthogonal comparisons of the ID groups show that ID-300 differs reliably from ID-0 and ID-100, $t\ (36) = 2.5$, $p < .02$.

Another measure of transfer is the comparison of shift performance with acquisition. The ID-300 group showed significant positive transfer, $t\ (7) = 4.5$, $p < .01$, whereas the ED-300 group showed significant negative transfer, $t\ (7) = 2.6$, $p < .05$. No other group showed either negative or positive transfer, $p > .05$.

The data of the study are plotted (see Fig. 1A) as backward learning curves (Hayes, 1953). In Fig. 1B is shown a family of computer-generated "stat-children." These theoretical curves show the relationship between the length of the presolution plateau and the probability of attending to the relevant dimension, $Po_{(1)}$ (Zeaman & House,

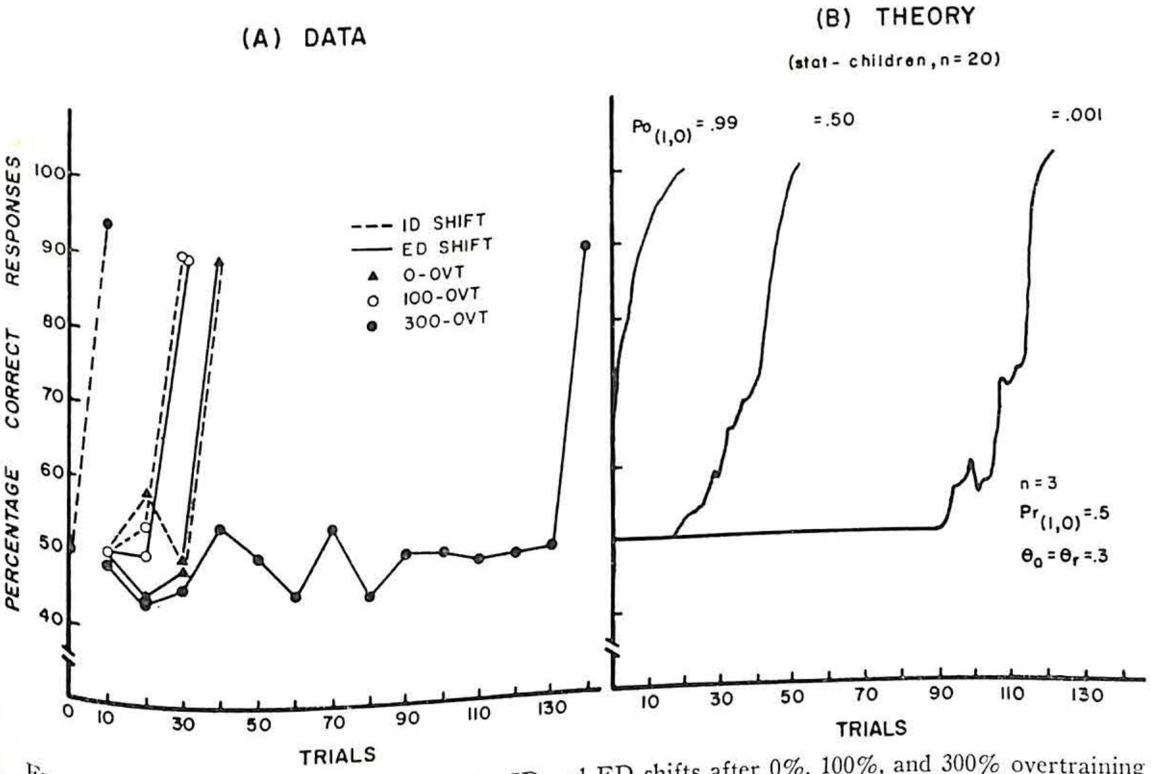


FIG. 1. (A) Backward learning curves for ID and ED shifts after 0%, 100%, and 300% overtraining; (B) theoretical curves showing the mean performances of three groups of "stat-children" under indicated parameter conditions (n = number of inferred dimensions competing for attention; $Po_{(1,0)}$ = probability of attending to relevant dimension at the start of the transfer problem; $Pr_{(1,0)}$ = conditional probability of choosing the positive cue at the start of training, given that attention has been directed to the relevant stimulus dimension; θ_o and θ_r = learning rate parameters for the attention and instrumental responses, respectively. The abscissa locus of each of the backward learning curves is that of the median S of a particular group.)

1963). This relationship between $Po_{(1)}$ and the presolution plateau implies that increasing the amount of overtraining should increase the length of the presolution plateau for the ED shift but reduce it for the ID shift. The agreement between the two sets of curves confirms the predictions of the model. By inspection, the postplateau slopes do not differ, and although the presolution plateaus of the data appear to be more variable than the theoretical plateaus, the former plateaus do not violate an assumption of stationarity, $t\ (15) = 1.3,\ p > .05$.

DISCUSSION

The Amount of Overtraining × Type of Shift interaction agrees with predictions of the models by Zeaman and House (1963) and Lovejoy (1966). The models assume that the probability of observing or attending to the relevant dimension approaches an asymptotic value more slowly than does the probability of approaching the correct cue. Thus with a weak criterion or a few overtraining trials, the relevant observing response may be relatively weak, although *S*s are consistently correct in their choices. With increasing amounts of overtraining, the relevant observing response becomes stronger and transfer to subsequent discriminations becomes greater. Consequently, when the relevant dimension in acquisition remains relevant in a second discrimination (ID shift), positive transfer is arranged and the second discrimination should be increasingly facilitated with increases in the amount of overtraining. When, however, the relevant dimension in acquisition is made irrelevant in a second discrimination (ED shift), negative transfer is arranged and the discrimination should be increasingly impaired with increases in overtraining. The performances of the ID-300 and ED-300 *S*s relative to the other ID and ED *S*s can be described accurately by the models.

Zeaman and House (1963) also predict specific details of the learning functions in transfer. The theoretical learning curves in Fig. 1B show that length of the presolution plateau varies inversely with the initial probability of observing the relevant dimension. Moreover, Zeaman and House have also shown that the manipulation of the learning rate parameters, θ_o and θ_r, does not produce a family of curves resembling those in Fig. 1B. With increasing amounts of overtraining, the initial probability of observing the relevant dimension at the beginning of transfer becomes higher for an ID shift and lower for an ED shift. Consequently, the plateau should become shorter as ID *S*s are given increasing amounts of overtraining, whereas increases in overtraining should lengthen the plateau for ED *S*s. An inspection of the functions in Fig. 1A and 1B indicates that the theoretical and empirical curves agree. No attempt has been made to estimate parameter values precisely. The values used to generate the "stat-children" have been used previously to compare theoretical and empirical functions (Shepp & Zeaman, 1966) and were chosen to approximate the form and order of magnitude of the empirical functions. Although Lovejoy (1966) has not presented these types of theoretical curves, the assumptions of his model clearly imply them.

The results of other shift studies in which the amount of overtraining has been varied (see Wolff, 1967) cannot be easily compared with the present results. In the present study, new stimuli of both the relevant and irrelevant dimensions were presented in transfer, allowing the direct assessment of attentional transfer. In contrast, most previous studies have retained in transfer one or more relevant cues that were presented in original learning and do not, therefore, permit a direct assessment of attentional transfer. However, two previous investigations have compared ID and ED shift performances as a function of the amount of overtraining. Eimas (1966) reports that ID shifts are facilitated by overtraining; the present results agree. In addition, Eimas reports that ED shifts also are facilitated by overtraining, a result not in agreement with the present study. The many procedural differences between the two studies preclude a close comparison; however, there are two methodological problems associated with Eimas' experiment that require a cautious interpretation of his results. First, several *S*s were dropped from the experiment for failure to maintain criterion during overtraining, a selection bias which would tend to favor the overtrained groups. Second, two of the ED shift subgroups were markedly inferior to all other groups in original learning. This uncorrected sampling bias makes the performances during transfer difficult to interpret.

Furth and Youniss (1964) report that an ID shift is superior to an ED shift following both criterion and overtraining procedures. Overtraining had no effect on either shift, and no Shift × Overtraining interaction was observed. Unfortunately, differential novelty effects made these results difficult to interpret. Novel relevant stimuli were added to the ID shifts, and novel irrelevant stimuli were added to the ED shifts. Since responses to novelty alone could have produced the effects observed by Furth and Youniss, the experiment does not serve as an adequate test of the theoretical models.

REFERENCES

Campione, J., Hyman, L., & Zeaman, D. Dimensional shifts and reversals in retardate discrimination learning. *Journal of Experimental Child Psychology*, 1965, **2**, 255-263.

Eimas, P. D. Effects of overtraining and age on intradimensional and extradimensional shifts in children. *Journal of Experimental Child Psychology*, 1966, **3**, 348-355.

Furth, H. G., & Youniss, J. Effect of overtraining on three discrimination shifts in children. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 290-293.

Goodwin, W. R., & Lawrence, D. H. The functional independence of two discrimination habits associated with a constant stimulus situation. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 437-443.

Hayes, K. J. The backward curve: A method for the study of learning. *Psychological Review*, 1953, **60**, 269-275.

House, B. J., & Zeaman, D. Reversal and nonreversal shifts in discrimination learning of retardates. *Journal of Psychology*, 1962, **63**, 444-451.

Lovejoy, E. Analysis of the overlearning reversal effect. *Psychological Review*, 1966, **73**, 87-103.

Shepp, B. E., & Eimas, P. D. Intradimensional and extradimensional shifts in the rat. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 357-361.

Shepp, B. E., & Zeaman, D. Discrimination learning of size and brightness by retardates. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 55-59.

Sutherland, N. S. Stimulus analyzing mechanisms. In, *Proceedings of a symposium on the mechanization of thought processes*. Vol. 2. London: Her Majesty's Stationery Office, 1959.

Wolff, J. L. Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin*, 1967, **68**, 369-408.

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.

(Received January 16, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 52–63

SUBJECTIVE AVERAGING OF LENGTH WITH SERIAL PRESENTATION[1]

DAVID J. WEISS[2] AND NORMAN H. ANDERSON

University of California, San Diego

The *S*s estimated the average of several lengths presented serially one at a time. In Exp. I, the judgment was made only at the end of the sequence. In Exp. II, *S* estimated a cumulative average as each new length was presented. The main phases of these experiments used sequences of six lengths. For the most part, each *S*'s data could be described by a subjective averaging model as tested in single-*S* analyses. There was a general recency effect, the later lengths in the sequence having greater influence. Recency was fairly uniform across *S*s with the end responding procedure of Exp. I, but large individual differences in the serial position curves appeared with the continuous responding procedure of Exp. II. In Exp. III, two hypotheses about the cause of recency were tested, but received little support. Functional measurement technique showed that subjective length differed from objective length, apparently by a constant error for each *S*. It was noted that the present methods could be applied to psychophysical scaling of other stimulus dimensions.

Suppose that you are shown several lines, one at a time, and asked to estimate their average length. Can your response be described as a mathematical average? If so, is it an arithmetic mean, geometric mean, weighted midpoint, or some other measure of central tendency? The present experiments were designed to consider this question. They test a subjective averaging model for length.

Averaging model.—The basic assumption of the model is that the response (R) at Serial Position N is simply a weighted sum:

$$R_N = \sum_{1}^{N} w_k s_k. \qquad [1]$$

Here w_k is the weight of the kth stimulus, and s_k is its scale value. For the present application, it is assumed that the w_k sum to unity across serial positions, so that Equation 1 is an arithmetic averaging model.

Two further restrictions are also made. First, the weight of any stimulus is assumed to depend only on serial position. Second, the scale value or subjective length of any stimulus is assumed to be constant, independent of the other stimuli in the sequence. Different *S*s, of course, will have different weights and scale values, but this is taken into account in applying the model.

Goodness of fit.—If mathematical correctness were taken as the standard, three sources of inaccuracy might arise. First, subjective length might differ from objective length. Second, the serial positions might receive unequal weighting, e.g., if weight were related to the recency of the memory trace. In either case, the subjective averages will generally be unequal.

Both the aforementioned sources of inaccuracy are allowed by the model. They correspond to values of w_k and s_k that differ from the arithmetically correct values. Neither of these inaccuracies, therefore, poses any immediate problem.

There is a third source of inaccuracy that would reflect adversely on the model as applied here. If the stimuli interact with one another, or with serial position, then the model will not hold in general. Such interactions might arise from psychophysical assimilation or contrast, or from other sorts of context effects.

It is important, therefore, to test the model, even while allowing for subjective values of the weight and value parameters that are specific to each *S*. Fortunately, testing is straightforward. If the stimuli are constructed from a factorial design,

---

[1] This work was supported by National Science Foundation Grants GB-3913 and GB-6666.

[2] Requests for reprints should be sent to David J. Weiss, Department of Psychology, University of California, San Diego, P. O. Box 109, La Jolla, California 92037.

analysis of variance provides a direct and powerful test of goodness of fit (Anderson, 1962a, 1964a).

Functional measurement.—If the model is validated, it may be used to scale both subjective length and the weights at each serial position. This application of functional measurement is illustrated for the present experimental design.

In the main body of the present data, there were just two lengths, 16 and 24 cm., at each serial position. The design yields estimates of their observed differential effect (D_k) at Serial Position k. The theoretical expression for D_k is:

$$D_k = w_k(s_{24} - s_{16}). \qquad [2]$$

Since the difference in scale values ($s_{24} - s_{16}$) is the same at each serial position, D_k is proportional to w_k. Hence, if each D_k is divided by their sum, these quotients add to unity and provide estimates of the w_k. These estimates define the serial position curve.

Subjective length may also be scaled using the model. At Serial Position N, the model implies that

$$\sum_1^N D_k = \sum_1^N w_k(s_{24} - s_{16}) = s_{24} - s_{16},$$

since the w_k sum to unity. Also, the grand mean response equals $(s_{24} + s_{16})/2$. The data thus provide estimates of both sum and difference of the scale values and hence of each scale value separately.

METHOD

The same basic length-averaging task was employed in all three experiments reported here. In Exp. I, S saw three or six lengths in sequence and estimated their average length at the end of the sequence after the last length had been presented. In Exp. II, the procedure was quite similar except that S estimated a running average, giving a judgment after each successive length was presented. In Exp. III, a procedure of simultaneous-serial presentation was introduced to test two hypotheses about the source of the recency effect observed in Exp. I and II.

Experiment I

Apparatus and procedure.—The line stimuli in this experiment were displayed horizontally on a 43×28 cm. screen using a rear projector. The projected lines were black, .5 cm. wide, and ranged from 14 to 26 cm. in length. The screen was 168 cm. in front of S, and the lines were displayed 28 cm. above the response panel. Each line was projected for 4 sec., with an interstimulus interval of 2 sec. during which the screen was blank. Normal room illumination was used.

Responses were made by the method of reproduction after the last line stimulus disappeared. By throwing a two-way, spring-loaded switch, S controlled a motor that adjusted the length of a looped horizontal white tape displayed in a panel 122 cm. in front of him.

On E's side of the response apparatus, a marker attached to the response tape passed along a meter stick to monitor S's response. Responses were read to the nearest millimeter by E, who sat concealed behind a screen by the side of the response apparatus. After each sequence, E reset the white tape to zero length.

Design.—In the main part of the experiment, there were six lengths in each sequence. At each of these six serial positions, the length could be long (24 cm.) or short (16 cm.). There are then 64 possible sequences, and they form a 2^6 factorial design.

Each regular daily session included a repetition of the instructions, 4 practice sequences, a half replication (32 sequences) of the basic design, plus 6 interspersed filler sequences. The practice and filler sequences included lengths of 14, 18, 22, and 26 cm., as well as 16 and 24 cm., and some of these sequences also served as end anchors. Although end effects did not seem likely with a continuous, unmarked scale, these anchor sequences were an extra precaution against distortion of the effective response scale.

After an initial practice day, each S was run in eight daily sessions, yielding four replications of the main design. This was followed by two sessions in which sequences of three lengths were judged. Analogous to the main design, this yielded a 2^3 design; each S was run through it four times each session.

Subjects and instructions.—The Ss were eight students who received $1.85 for each of the 11 sessions. They were told to estimate the average length of the six (or three) lengths in each sequence, and some preliminary sequences were used to ensure that they understood the task.

Experiment II

General procedure was similar to that of Exp. I, except that S responded after each length of the sequence, giving the running average of the lengths he had seen so far. In addition, the stimulus display and response panel were somewhat different. Experiment II was run before Exp. I, but is presented second for expositional simplicity.

Apparatus and procedure.—The line stimuli in this experiment were sticks, 1 cm. in diameter, painted black, and of the same lengths as used in Exp. I. They were presented singly between guides on the table 60 cm. in front of *S* and centered about 25 cm. left of the zero region on the response apparatus.

Responses were made with the stick present. Responding was self-paced and generally required 4–5 sec. per length. A complete sequence of six lengths required about 1 min.

The response apparatus consisted of a meter stick surmounted by a brass rod that held a sliding indicator about 5 cm. above the table. On the side facing *S*, the meter stick was covered with black felt, except for the rightmost 20 cm., which was white. The black-white boundary served as the zero reference from which *S* estimated the average length. Within each sequence, the response indicator remained at the last response until *S* made his new estimate. At the end of each sequence, *S* returned the indicator to the zero position.

The *E* sat on the other side of the table from *S*, presented each stick in turn, and recorded the response by reading the meter stick to the nearest millimeter. Except for the one present stimulus, the sticks were concealed from *S* behind a small blind.

Design: Phase 1.—In the main part of the experiment, there were six lengths in each sequence. At each of these serial positions, the lengths could be long (24 cm.) or short (16 cm.). As in Exp. I, there are thus 64 possible sequences and they form a 2^6 factorial design.

Each regular daily session included a repetition of the instructions, four practice sequences, a half replication (32 sequences) of the 2^6 design, plus six interspersed filler sequences. Practice and fillers were the same as in Exp. I. Each *S* served one practice and eight regular sessions, yielding four replications of the basic design.

Design: Phase 2.—In this phase, 10 lengths were used in each sequence in order to obtain serial curves for longer sequences. Procedure was the same as in Phase 1: a 2^6 design was again employed. To compress 10 serial positions into six factors, each of the first four pairs of serial positions was treated as a factor, each with the same two levels. One level consisted of 14 cm. followed by 18 cm.; the other level consisted of 22 cm. followed by 26 cm. The levels of the last two factors, Serial Positions 9 and 10, were 16 cm. and 24 cm., as before. Each *S* went through the complete design twice at the rate of one-quarter replication per day.

Subjects and instructions.—Undergraduate *S*s received $1.85 per session, except for *S*4, a department secretary. An initial squad of four *S*s was run through both phases of the experiment. The data supported the model, but there were large individual differences in the serial position curves. To get more information on the distribution of individual differences, *S*5–*S*10 were then run through the Phase 1 design.

Running average instructions were employed. The *S*s were told to simply estimate the length of the first length, then the average of the first two lengths, etc. This pointer position was then to be changed to incorporate each new length in the cumulative average. A movement of the pointer was required at each stimulus presentation; however, *S* was told that he could move it very slightly, or away and back.

Experiment III

A simultaneous-serial presentation procedure was employed to test two hypotheses about the recency effect. The main procedural change was to present several lengths simultaneously in the first serial position. Thereafter only one length at a time was presented, and *S* responded after each presentation with a running average, as in Exp. II.

Apparatus and procedure.—The response apparatus was the same as in Exp. I. The stimuli were sticks, as in Exp. II, ranging 10–28 cm. in 2-cm. steps. Stimulus presentation was the same as in Exp. II, except that from 1 to 5 lengths might be shown simultaneously on the first serial presentation of a sequence. When several lengths were shown together, they rested on separate guides parallel to the response panel, with the sticks placed haphazardly among themselves. Responding was self-paced and generally required about 10 sec. per response for the naive *S*s of this experiment.

Design.—There were seven basic conditions, as shown in Table 3. The first number in the condition designation is the number of lengths presented together in the first serial position. Thus, in Cond. 4-1-1, four lengths were shown together initially; then the next two serial positions had one length each, for a total of six lengths.

Six different pairs of lengths, each pair differing by 8 cm., were used for the first five conditions listed in Table 3. Length pairs were balanced over serial positions with a 6 × 6 Latin square, as in Anderson (1964b). As a consequence, the lengths in the initial presentation were all different for the first five conditions.

For the last two conditions listed in Table 3, the four or five lengths presented in the initial position were all equal. It was thought that this would produce greater confidence in the estimated average than when the first four or five lengths were all different, thereby reducing the recency effect.

Each *S* judged a total of 64 sequences, from 4 to 16 in each of the seven conditions. As in the previous experiments, these were constructed from factorial designs in order to allow estimation of the weights associated with each serial position. For the longer sequences, this required use of fractional replication (Cochran & Cox, 1957).

Subjects and instructions.—The 24 undergraduate *S*s were each paid $1.00 plus class credit for the 1.5-hr. session. Instructions were similar to those used in the first two experiments.

RESULTS

The main tests of the model, in Exp. I and II, were made by applying analysis of variance separately to the data of each single *S*. The theoretical basis for this analysis has been noted previously, and here it need be kept in mind only that the model implies that the interaction terms are zero. The within-replicates variability was used as the error term in these single-*S* analyses. Certain group analyses were also made for various purposes. In these, of course, the error term is the appropriate *S* interaction.

There is a difficulty that is partly an embarrassment in reporting these analyses because of the multitude of interactions. With six serial positions, a complete factorial yields 57 interaction terms, and these are difficult to present in simple form. Moreover, with so many tests, just about three interactions might be expected to be significant at the .05 level by chance alone. It is desirable to avoid testing so many interactions if that can reasonably be done (Anderson, 1968). A few remarks on this question, written in light of subsequent results, may conveniently be included at this point.

There seem to be three main ways to reduce the number of interactions tested. One method is to employ confounding procedures (Cochran & Cox, 1957). For instance, a 2^6 design in $\frac{1}{8}$ replicate leaves just 1 *df* for a test of pooled interactions. It also reduces the number of stimulus sequences from 64 to 8, a feature that is useful in many experimental applications (e.g., Anderson, 1964b).

A second method is to apply partial analysis (Anderson, 1968). With partial analysis, high-order interactions are not tested when there is reasonable ground to expect that they will be negligible. Related to this is the third method, which focuses on particular interactions or other tests in which there is specific reason to expect discrepancies from the model to be located. For example, contrast or other patterning effects might be expected to appear largely in the two-way interactions of successive serial positions.

Since this was an initial investigation of length averaging and since a previous experiment on loudness averaging (Parducci, Thaler, & Anderson, 1968) had raised some doubt about the model, it was decided to employ a complete design and analysis. Fortunately, the general pattern of the results supports the use of the aforementioned methods to reduce the number of interactions tested. In fact, confounding formed an essential part of the design in Exp. III.

Experiment I

In this experiment, the response was made only at the end of the sequence. Two aspects of these data are of interest: first, the test of the averaging model; second, the serial position curves.

Goodness of fit.—A summary of the tests of the model is given in Table 1. The tabulations include all separate interactions that were significant for each *S*, together with an overall *F* ratio based on all interactions pooled. The main data, for sequences of six lengths, are in the left half of the table. Since the 2^6 design has 57 interaction terms, an average of 3 separate interactions would be expected to be significant by chance alone. On such a rough assessment, *S*1

TABLE 1

SUMMARY TEST OF FIT, EXP. I

S	Six-length sequences				Three-length sequences			
	Interactions[a]	MS_e	Interaction *F* ratio[b]	r^c	Interactions[a]	MS_e	Interaction *F* ratio[d]	r^c
1	12, 23, 25, 26, 135, 145, 146, 345, 1256, 1356, 12346	.95	2.11*	.910	123	.70	3.88*	.981
2	12345	1.59	.82	.956	None	.98	.51	.996
3	23456	1.95	.90	.951	None	1.14	.26	.998
4	12	.95	.81	.963	123	.41	11.66*	.974
5	12, 34, 46, 234, 2356	2.48	1.14	.923	None	2.26	.13	.998
6	25, 36, 1246, 12356	2.64	1.25	.902	None	1.76	.44	.995
7	None	1.79	.63	.953	123	.19	3.50*	.995
8	56, 13456	5.75	.68	.944	123	.84	3.87*	.992

[a] Code numbers indicate significant interactions; each digit represents corresponding serial position.
[b] *F* ratio for pooled interactions, $df = 57$ and 192.
[c] Correlation (observed, predicted).
[d] *F* ratio for pooled interactions, $df = 4$ and 56.

clearly deviates from the model, showing 11 separate interactions, as well as a significant pooled interaction. The other seven *S*s appear to satisfy the test of fit reasonably well.

Inspection of Table 1 fails to show any clear pattern among the significant interactions; e.g., each serial position is represented with nearly equal frequency. Inspection of the data tables for the two-way interactions also failed to show any pattern. However, it was possible to localize the discrepancies for *S*1 in a single sequence. Inspection of the raw data showed an unusually high response to that sequence containing the longer length at all six serial positions. Subtracting a constant, 3.9 cm., from the response to each replication of this sequence reduced the number of significant interactions from 11 to 2, about what would be expected by chance. No other unusual responses were found, and the cause of this one discrepancy is unknown.

Deviations from the model that were consistent across *S*s would tend to show up in the overall test. In this group analysis, however, only two high-order interactions were significant, no more than would be expected by chance. This situation remained unchanged when the pooled within-*S*s error on 1,536 *df* was used for all the tests. These results suggest that what real discrepancies there are from the model are idiosyncratic. They may present difficulties, but they do not reflect any uniform preceptual effect.

The right half of Table 1 shows the summary for the last phase of the experiment, in which each sequence had three lengths. No two-way interactions were significant, but four *S*s show a significant 123 interaction. However, it took a different form in each case and did not approach significance in the group analysis, even when tested against the pooled within-*S* error on 448 *df*.

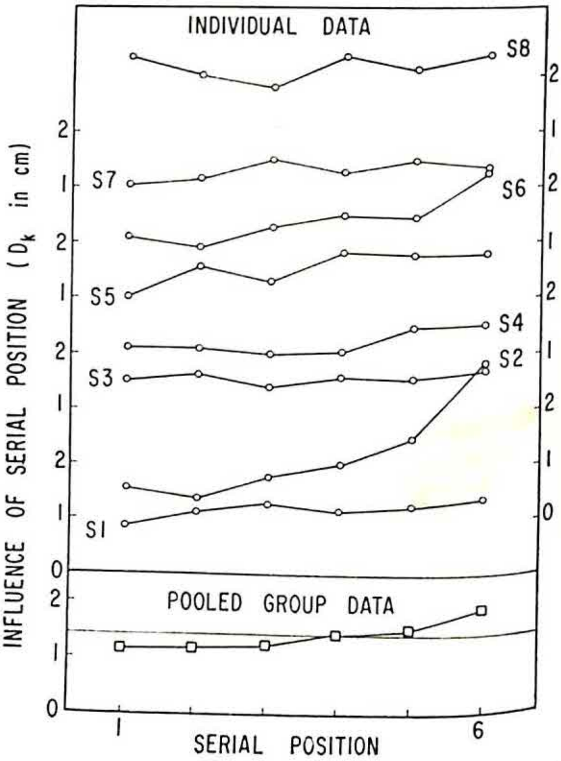


FIG. 1. Serial position curves, Exp. I. (Individual curves displaced upward, as indicated by repeating ordinates.)

The seriousness of the interactions depends heavily on their magnitudes. It is not feasible to report these in detail, but two data columns in Table 1 have relevant information. The first is the overall *F* ratio, already noted, which gives the magnitude of the total interaction component relative to the error variability. These are all near unity, except for *S*1. The second is the correlation between the observed mean response and that predicted by the model. These correlations are, in effect, the ratios of the *SS* for main effects to the total *SS* for systematic sources. Thus, the degree to which they fall short of unity is an index of the relative size of the pooled interactions. Although these correlations seem satisfactorily high, they must be interpreted with due care. In particular, that smaller correlations are obtained for six lengths than for three lengths is almost entirely artifactual, arising from the more numerous interaction terms in the former designs.

Serial position curves.—These curves, in Fig. 1, show the differential effect of the 24- and 16-cm. lengths as a function of serial position. With one exception, the curves are fairly similar in shape, showing a mild recency effect. For graphical clarity, the curves are separated vertically, as indicated in the graph.

The calculation and interpretation of these serial curves require a brief comment. The entry for Serial Position k is the difference, D_k, between two marginal means: for all those sequences that had the 24-cm. length at Position k, and for all those sequences that had the 16-cm. length at Position k. If the model is correct, then this observed difference is an estimate of $w_k(s_{24} - s_{16})$, where w_k is the weight for Position k and s_{24} and s_{16} are the subjective lengths of the two stimuli. If the difference in subjective lengths is indeed constant across serial position (see Exp. II), then

TABLE 2

SUMMARY TEST OF FIT: SIX-LENGTH SEQUENCES, EXP. II

Interaction	*S*									
	1	2	3	4	5	6	7	8	9	10
Serial Position 6										
12			4.49			5.89	6.99			5.05
34			4.19				4.92			
56	4.85									
134					6.48					
345					5.94					
1235					4.40					
1245		4.24								
1346									4.82	
12345				7.18						
12346		10.49	4.79							
13456				4.48						
Main[a]	632.50	185.92	223.77	38.01	158.18	86.37	175.22	219.52	164.72	64.85
MS_e (192 df)	.51	.50	.47	2.57	.85	1.77	.63	.83	.57	1.58
Serial Position 5										
12			6.82			6.67	8.90			
23			4.68							
34							4.70			
45				7.73						
134		3.89			6.12					
234							5.58			
345					6.26		4.85			
1235					6.61					
12345					6.75	4.70	4.44			
Main[a]	666.06	227.04	343.48	62.08	229.26	142.54	236.48	316.85	209.04	99.58
MS_e (224 df)	.57	.60	.41	2.63	.80	1.45	.61	.68	.57	1.45
Serial Position 4										
12			10.68			5.72	8.69			7.96
13			5.53							
23		4.40	5.53							
24					5.31		8.88			
123		4.01								
124						10.21			4.40	
134					4.86				11.23	
Main[a]	1044.13	343.15	578.05	105.13	356.73	214.24	367.90	368.73	365.60	162.68
MS_e (240 df)	.47	.61	.36	2.37	.76	1.43	.59	.74	.49	1.38
Serial Position 3										
12			12.75			5.70	11.30		3.95	
13			10.68							
23			5.26	6.61	4.86					
123									24.27	
Main[a]	1331.20	578.73	1015.33	287.50	596.50	355.50	602.67	627.37	704.40	290.43
MS_e (248 df)	.51	.66	.38	2.03	.76	1.36	.61	.65	.47	1.44
Serial Position 2										
12			23.00	13.67		4.70	14.61			3.93
Main[a]	2153.33	1213.73	2075.86	772.48	1241.37	731.95	1407.93	870.68	1845.90	630.59
MS_e (252 df)	.59	.71	.42	1.53	.76	1.50	.59	.76	.42	1.52
Serial Position 1										
MS_e (254 df)	.77	.85	.49	.98	1.00	1.35	.46	.61	.50	1.78

[a] Mean F ratio for main effects.

the curves of Fig. 1 are proportional to the values of w_k. In other words, these serial curves are the weight curves.

The serial curves for the sequences of three lengths are not shown here, but they also showed a mild recency. Moreover, each *S* showed about the same curve shape for three lengths as for six. This individual consistency in serial position curves was also found in Exp. II.

Experiment II

In this experiment, *S* responded to each successive length in the sequence. The analysis of the terminal response is exactly the same as in Exp. I. In addition, the data at each earlier serial position may be analyzed in the same way. The successive responses in a sequence are not independent, of course, and neither are the analyses. Nevertheless, the additional responses do give useful information.

Goodness of fit: Six lengths.—An overview of the tests of the model is given in Table 2. All interaction *F* ratios that were significant for each *S* for each serial position are listed. For comparative purposes, the mean *F* ratio for the main effects at each serial position is also included.

The upper section of Table 2 summarizes the tests of the response at Serial Position 6. These entries are comparable to the corresponding analysis of Exp. I summarized in Table 1. On a gross count, the two experiments show about the same number of significant interactions. In Exp. II, however, there is a marked tendency toward significant 12 interactions, between Serial Positions 1 and 2. This same tendency is found in the analysis of each of the previous responses. In addition, *S*3, *S*7, and perhaps *S*5 show more significant interactions than would be expected by chance. Indeed, *S*3 and *S*7 show a significant 12 interaction at each response.

All told, the 12 interaction accounts for nearly a third of the significant *F* ratios. The initial response is just a reproduction of the first length and presumably presents no difficulty. Accordingly, the source of the interaction would seem to be at Serial Position 2, the first position at which *S* is required to average. As already noted, successive responses are not independent, and it is straightforward to show that an interaction induced at one position tends to reappear automatically at subsequent positions. Presumably, therefore, the 12 interactions at the later positions reflect a discrepancy perpetuated from Serial Position 2.

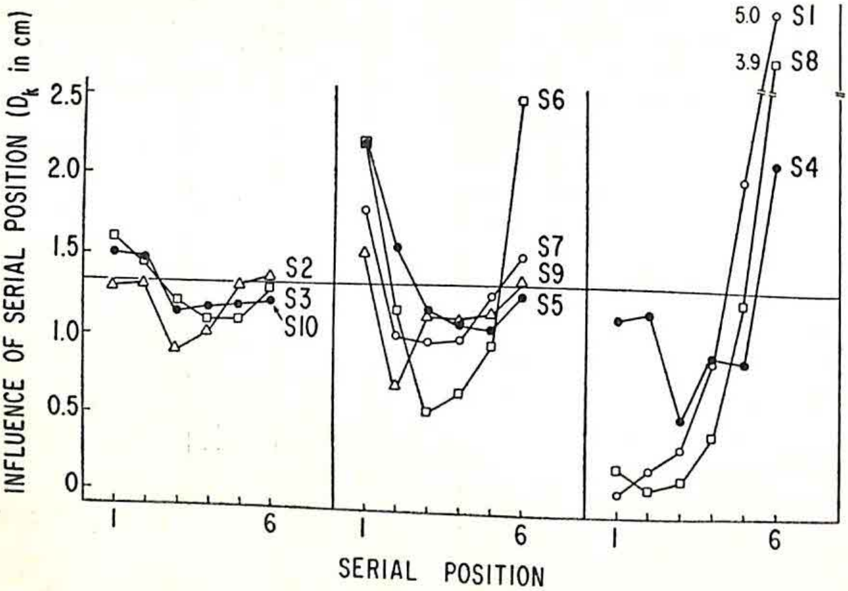


FIG. 2. Serial position curves for each *S*, Exp. II. (Curves based on data from response at Serial Position 6.)

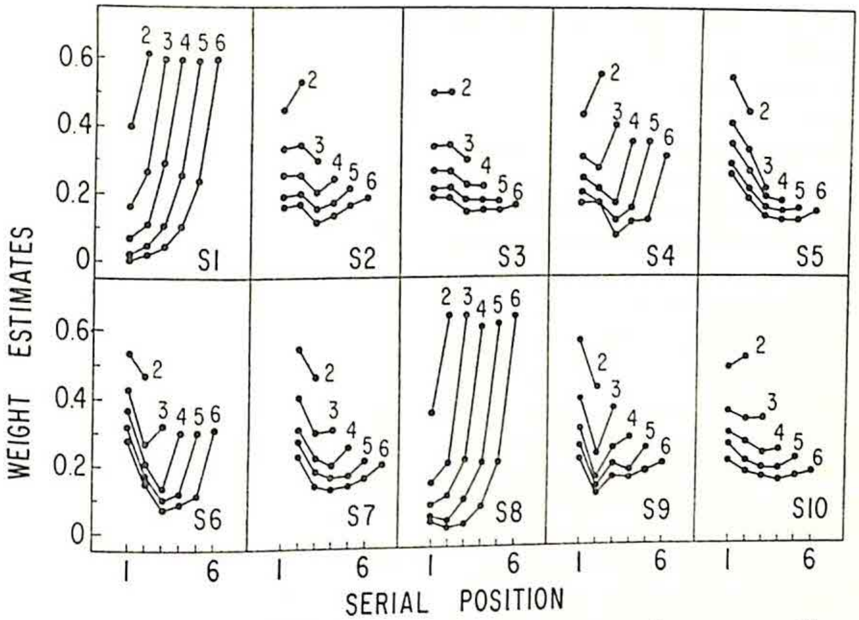


FIG. 3. Serial position curves for each *S*, for each successive response, Exp. II. (Numbers by curves in each panel index successive responses within sequence.)

Other than the 12 interaction, not much pattern is evident in Table 2. The mean total of significant interactions is 5.9 per *S*, only slightly greater than would be expected by chance. Even for the 12 interaction at Serial Position 2, there was no consistent pattern across *S*s. It had a different shape for different *S*s and was not significant in the group analysis.

Response variability was roughly constant across serial position. This can be seen by inspection of the error mean squares of Table 2. Despite considerable individual differences, each *S* shows about the same variability in response at each serial position. The detailed analyses of the successive responses for the sequences with 10 lengths showed a similar pattern.

Serial curves: Six lengths.—The serial curves calculated from the terminal response are shown in Fig. 2. For graphical clarity, *S*s have been grouped according to curve shape, which varies considerably. Some show extreme recency (right panel), some show relatively flat serial curves (left panel), and others show both primacy and recency components (center panel). This variety in curve shape is markedly greater than for Exp. I (Fig. 1), in which only a terminal response was made.

Serial curves may also be obtained from the earlier responses and these are shown in Fig. 3. These plots are in terms of *w* values in order to facilitate comparison of the several curves for each *S*. As can be seen, each *S* tends to have a stable characteristic curve shape.

Scale values: Six lengths.—The scaling procedure described in the introduction provides estimates of the subjective lengths at each serial position. These are shown

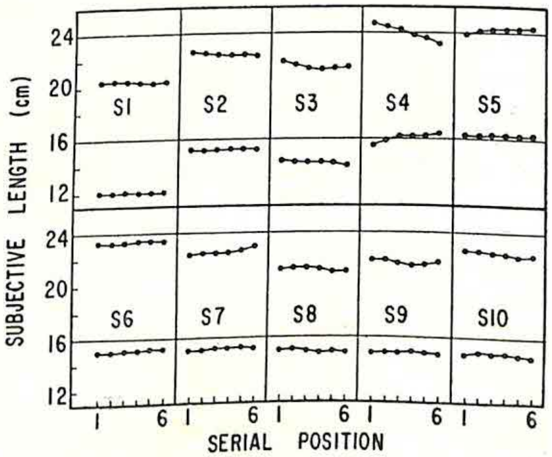


FIG. 4. Subjective length estimated from response at each serial position, Exp. II. (The two curves for each *S* give subjective lengths of long, 24-cm., and short, 16-cm., lengths.)

in Fig. 4. For nearly every *S*, subjective length is constant over serial position. This indicates that there are no within-sequence adaptation effects.

Figure 4 shows marked individual differences in the subjective lengths. These individual differences are real enough since the first point on each curve is nothing but a direct estimate of the first length in the sequence. This fact, it may be emphasized, helps validate the scaling procedure. Curiously enough, the difference between the two subjective lengths is fairly close to its arithmetically correct value of 8 cm. for every *S*. The individual differences in subjective length can, therefore, be considered as constant errors.

Serial curves: Ten lengths.—The main function of these data was to yield serial curves for longer sequences, and only the terminal response is considered here. The tests of fit showed about the same general picture as in the main phase. Of the 228 interactions, only 14 reached significance. However, *S*3 showed an array of discrepancies similar to that noted previously, and *S*4 appeared to have some possibly real discrepancy from the model. As before, these discrepancies tended to be relatively small in magnitude, and inspection of the data failed to show any pattern among them.

The serial curves are shown in Fig. 5. Although there are 10 serial positions, there are only 6 points on the curves since the design used pairwise confounding over the first 8 serial positions. The most important aspect of these data is that *S*s show the same shape curve with 10 lengths as with 6 (compare Fig. 5 with Fig. 2). This individual consistency in shape of serial curve confirms that found in Exp. I.

Averaging accuracy.—In the length-averaging task, there is an objective criterion of accuracy to which the behavior may be compared. With serial presentation, of course, the recency effect guarantees inaccuracy. Nevertheless, the balance in the stimulus design allows a comparison of objective length and subjective length independent of recency. This is obtained by averaging over all sequences with the same frequency distribution of lengths regardless of their position. Figure 6 plots subjective average as a function of objective average for the sequences of six lengths. The second point on the curves is the mean response to all six sequences that had just one short length, etc. With end responding (Exp. I), *S*s are quite accurate. With continuous responding (Exp. II), marked underestimation occurs. This constant error parallels that of Fig. 4. Amount of constant error varies from *S* to *S*, but to the degree that the behavior follows the averaging model, each *S*'s curve would be parallel to the group curve.

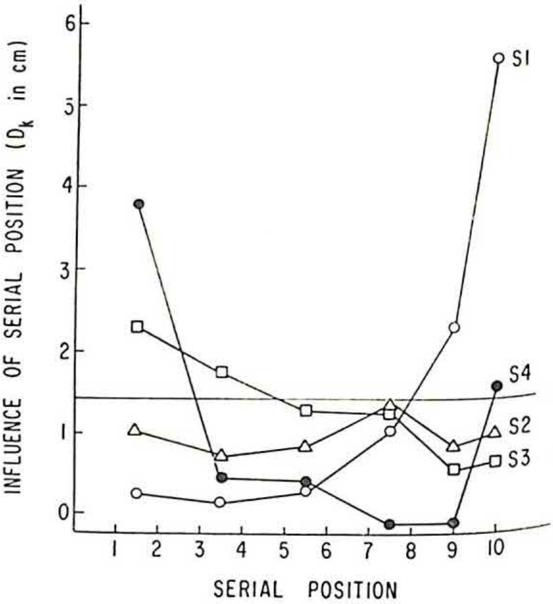


FIG. 5. Serial position curves for the four *S*s in Phase 2 of Exp. II. (Compare with Fig. 2.)

Experiment III

A method of simultaneous-serial presentation was adopted in this experiment to test two hypotheses about the recency effect observed in Exp. I and II. The seven types of sequences are shown in the left column of Table 3. In each row, the first number in the sequence type is the number of lengths presented simultaneously in the initial position of the sequence; thereafter only one length was presented at a time.

The entries in Table 3 are the weights estimated from the terminal response. The weight for the first serial position, which represents several stimuli, was reduced to a per stimulus basis as indicated in the table. Two principal features of Table 3 merit comment. First, w_6 is the largest entry in each row, a sizable recency effect. Second, the prorated weight of a member of the initial set is roughly constant for sets of more than one length.

To understand the two hypotheses about the recency effect, S's task should be kept clearly in mind. At any step in the sequence, both the new stimulus and S's previous estimate are in view. All S need do is average them, and Exp. I and II indicate that he does just that. Since both relevant stimuli are visibly present, the recency presumably does not arise from any misperceptions or memory changes in the scale values. Accordingly, the problem is why S overweights the new stimulus and underweights the previous estimate.

Now the weight associated with the previous estimate is, in fact, directly related to the number of previous stimuli. This weight, therefore, may be interpreted as a subjective N. The first hypothesis is that S does not keep in mind the exact number of previous stimuli, but that this subjective N undergoes a temporal decay. If this is true, then the recency should be less for the shorter sequences, in which several stimuli are given simultaneously at Serial Position 1.

This hypothesis is tested by the first five sequence types of Table 3. It implies that w_6 should decrease as the size of the initial simultaneous set increases. Inspection of the data shows no trend, and the statistical test was nonsignificant, F (4, 92) = 1.99. Similar data from the fifth and fourth responses also showed no effect of size of initial set. In short, there is no support for the first hypothesis.

TABLE 3

EFFECTS OF EACH SERIAL POSITION ON FINAL RESPONSE: WEIGHT ESTIMATES, EXP. III

Sequence type	w_1	w_2	w_3	w_4	w_5	w_6
1-1-1-1-1-1	.21	.12	.13	.13	.16	.26
2-1-1-1-1	(.14)	(.14)	.14	.12	.14	.31
3-1-1-1	(.13)	(.13)	(.13)	.16	.19	.24
4-1-1	(.14)	(.14)	(.14)	(.14)	.13	.33
5-1	(.15)	(.15)	(.15)	(.15)	(.15)	.26
4-1-1 (same)	(.16)	(.16)	(.16)	(.16)	.15	.21
5-1 (same)	(.15)	(.15)	(.15)	(.15)	(.15)	.25

Note.—Entries in parentheses are weights of initial set on a per stimulus basis. Initial simultaneous set in last two rows had all equal lengths. Sum of w values in some rows differs from 1.00 because of rounding error.

The second hypothesis states that the overweighting of the present stimulus is caused by lack of confidence in the previous response. The last four rows of Table 3 test this hypothesis. In the last two rows, the four or five lengths of the initial set were all equal; in contrast, they were all different for the next-to-last two rows. The confidence interpretation would imply less recency for the "same" than for the "different" condition.

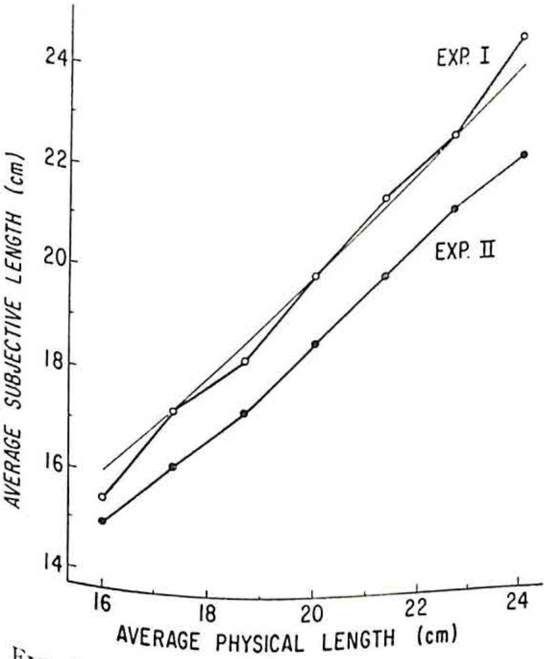


FIG. 6. Subjective average as a function of objective average, sequences of six lengths, with recency effect averaged out.

Some slight support for the confidence hypothesis was obtained from Sequences 4-1-1 since the difference between the estimates of .33 and .21 for w_6 was significant, F (1, 23) = 8.43. However, the corresponding test on w_5 estimated from the previous response was not significant. Moreover, the data for Sequences 5-1 in Table 3 show no effect.

DISCUSSION

This discussion takes up two main topics. The first is the problem of psychophysical integration, here exemplified by the serial averaging task. On the whole, despite certain discrepancies noted previously, the model seems to have done reasonably well for subjective length averaging. This result is not trivial. It is true that Ss were instructed to "average," but they were under no constraint about how they averaged. There are many measures of central tendency, and each S was free to choose his own. This averaging rule could even be variable, depending on the particular sequence of lengths. It has considerable interest, therefore, that the behavior was well described as a weighted arithmetic mean, with weights dependent only on serial position. Moreover, this result is important in the interpretation of the serial position curves.

In this connection, the predominant recency observed in each of the experiments needs consideration. When a response was required only at the end of the sequence, individual differences were relatively small (Fig. 1) and all Ss showed moderate recency. But when a response was required at each successive serial position, individual differences were extreme (Fig. 2, 3, and 5). Although these two experiments differed in procedural details and so are not strictly comparable, the difference between them is sufficiently strong to deserve attention.

Memory storage requirements might be related to the difference in range of recency in the two main experiments. With continuous responding, no memory of the individual previous stimuli is required since their average is represented visibly in the current position of the response indicator. Even so, *S* still must assign weights in some manner, and this requires a revision at each successive serial position of the relative weight to be assigned to the cumulated average. The large individual differences obtained with continuous responding would then presumably be cognitive, rather than perceptual, reflecting different weighting strategies. This is consistent with the within-*S* consistency in curve shape. With end responding, on the other hand, *S* could simply store the several subjective lengths more or less separately and make only one overall integrative response, thereby avoiding the successive weighting revisions. Task variations based on intermittent responding and component recall might help pin down this problem.

In any case, the recency itself remains to be explained. For continuous responding, Exp. III indicates that recency does not arise from decay of subjective *N* with time or serial position. The second recency hypothesis, that lack of confidence in the previous judgment produced recency, received only weak, equivocal support. In view of the large individual differences with continuous responding, it might be preferable to use an end responding procedure in future tests of this kind.

Whatever the cause of the recency, it appears regularly in serial averaging of simple perceptual stimuli. Besides the present length dimension, recency has been found with loudness (Parducci et al., 1968) and lifted weights (Anderson, 1967; Anderson & Jacobson, 1968), as well as with numbers (Anderson, 1964b). These stimulus classes do not, however, give uniform support to the averaging model. Loudness averaging, in particular, showed a sizable discrepancy. In weight averaging, the model did well with six weights, but a possible discrepancy appeared with three weights. This is consistent with the present indications of interactions at the second and third serial positions.

The second topic for discussion is functional measurement and its application to psychophysical scaling. Functional measurement provides a basis for scaling both stimulus and response dimensions at the same time. The essential ideas were given in Anderson (1962a, 1962b); recent applications are given by Shanteau and Anderson (1969) and Anderson and Jacobson (1968). Psychophysical averaging involves not only the subjective value of the stimulus, but also its subjective weight. The two are not always separable, but the continuous responding procedure has technical interest because it provides a basis for measuring both value and weight at each serial position.

A key feature of functional measurement is the use of an integration task, in which several stimuli are to be combined into a single judgment. The quantities that *S* combines are, of course, the subjective values of the stimuli. If the integration rule is simple, then the stimulus values may appear fairly directly in the response. This is one reason for attempting to develop an averaging model: in its simplest applications, the marginal response means of the stimulus design are equivalent, up to a linear transformation, to the subjective values of the stimuli. This remains true, of course, when more than two stimulus values are used in each factor of the design, a desirable property in determining the psychophysical law relating subjective and objective stimulus values.

Garner (1954a, 1954b), Garner and Creelman (1967), and Treisman (1964) have pointed out the perplexities in the interpretation of psychophysical scales based on direct, numerical response methods. For the most part, these methods obtain judgments of single stimuli and yield results of the form $R = F(S)$, where R is the observed response, S is the physical value of the stimulus, and F is the presumptive psychophysical law. Ordinarily, S is known. Hence, if R could be taken at face value, F would be completely determined, and the problem would be solved. But if R is simply assumed to be valid, then any observed relationship is arbitrary. As both Garner and Treisman have emphasized, numerical response methods typically take R at face value and provide no means for assessing that assumption. Similarly, Luce and Galanter (1963) have noted that numerical biases may vitiate the use of numerical response measures.

Numerical response measures can often be considered no more than ordinal scales. The proper response scale, R_T, if it exists, will then be some monotone transformation, $R_T = \mathfrak{M}(R)$, of the observed response (Anderson, 1962b). If $R = F(S)$ is observed, then the proper psychophysical law will be $\mathfrak{M}(F)$ since $R_T = \mathfrak{M}(R) = \mathfrak{M}(F(S))$. In the absence of any validation criteria, all continuous, strictly monotone transformations of F would have equal claim to be the psychophysical law.

The methods proposed here avoid this indeterminacy. The integration task provides a basis for determining a relation of the form $R = f(s_1, s_2, \ldots, s_N)$, where s_k is the subjective value of the kth stimulus. Because R is based on more than a single stimulus, constraints are available that allow application of a monotone rescaling procedure for the response dimension (Anderson, 1962b). In the present case, f was a weighted average, and the observed R satisfied the model so that no response rescaling was needed. Given the validity of the response, the subjective values of the stimuli could be derived fairly simply, in part because the integration rule was simple, in part because the stimulus combinations were constructed according to a factorial design. In this formulation, the psychophysical law is expressed as $s_k = F(S_k)$ and would appear as a by-product of the investigation. To determine this function would require using more values of S_k, of course, but in principle this is no problem.

The validity of functional measurement will depend on getting the same scales with different integration tasks. This follows Seward's (1955) requirement of constancy of the intervening variable, as well as the emphasis on converging operations of Garner, Hake, and Eriksen (1956). For this reason, it is desirable to develop alternative integration tasks. Simultaneous presentation has certain advantages over serial presentation. For length, both averaging and summing tasks may be used, with many variations of serial and simultaneous presentation. Few stimulus dimensions have this flexibility, but an averaging task appears feasible with almost any dimension. Indeed, traditional bisection judgments may be considered as averages of two stimuli, analogous to the present response at the second serial position. Although the present approach depends primarily on gross stimulus differences and a numerical response, choice data based on fine stimulus differences would also be of interest. There are various ways to begin the analysis of choice data, but the discriminant function approach of Rodwan and Hake (1964) has special interest and potential.

REFERENCES

Anderson, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817–818. (a)

Anderson, N. H. On the quantification of Miller's conflict theory. *Psychological Review*, 1962, **69**, 400–414. (b)

Anderson, N. H. Note on weighted sum and linear operator models. *Psychonomic Science*, 1964, **1**, 189–190. (a)

Anderson, N. H. Test of a model for number-averaging behavior. *Psychonomic Science*, 1964, **1**, 191–192. (b)

Anderson, N. H. Application of a weighted average model to a psychophysical averaging task. *Psychonomic Science*, 1967, **8**, 227–228.

Anderson, N. H. Partial analysis of high-way factorial designs. *Behavior Research Methods and Instrumentation*, 1968, **1**, 2–7.

Anderson, N. H., & Jacobson, A. Further data on a weighted average model for judgment in a lifted weight task. *Perception and Psychophysics*, 1968, **4**, 81–84.

Cochran, W. G., & Cox, G. M. *Experimental designs*. (2nd ed.) New York: Wiley, 1957.

Garner, W. R. Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, 1954, **48**, 218–224. (a)

Garner, W. R. A technique and a scale for loudness measurement. *Journal of the Acoustical Society of America*, 1954, **26**, 73–88. (b)

Garner, W. R., & Creelman, C. D. Problems and methods of psychological scaling. In H. Helson & W. Bevan (Eds.), *Contemporary approaches to psychology*. Princeton, N. J.: van Nostrand, 1967.

Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychological Review*, 1956, **63**, 149–159.

Luce, R. D., & Galanter, E. Psychophysical scaling. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.

Parducci, A., Thaler, H., & Anderson, N. H. Stimulus averaging and the context for judgment. *Perception and Psychophysics*, 1968, **3**, 145–150.

Rodwan, A. S., & Hake, H. W. The discriminant function as a model for perception. *American Journal of Psychology*, 1964, **77**, 380–392.

Seward, J. P. The constancy of the *I-V*: A critique of intervening variables. *Psychological Review*, 1955, **62**, 155–168.

Shanteau, J. C., & Anderson, N. H. Test of a conflict model for preference judgment. *Journal of Mathematical Psychology*, 1969, **6**, 312–325.

Treisman, M. Sensory scaling and the psychophysical law. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 11–22.

(Received January 17, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 64–69

SUCCESSIVE CONTRAST EFFECTS AS A FUNCTION OF TYPE AND MAGNITUDE OF REWARD[1]

I-NING HUANG[2]

University of Texas at Austin

Six groups of nine rats each were trained, one trial a day, in a straight runway in order to investigate the effects of a Phase 1 reward condition on *S*'s subsequent behavior in terms of its influence on reward shift (Phase 2) and reacquisition (Phase 3) following either a different reward quality, a different reward magnitude, or an interpolated training with nonreward. A marked depression effect was obtained when the shift was to a qualitatively different reward (e.g., food to water) with a similar reinforcing potency. The reacquisition rate was faster for all conditions of interpolated training when reward was the same in Phase 1 and Phase 3 rather than when it was different. The findings provided support to the sequential hypothesis.

In the usual instrumental learning situation, a motivated *S* not only learns the instrumental response, but also learns about the goal event, i.e., the quality and the quantity of the reward. This learning about goal events has been variously termed the anticipatory goal response (r_G, Hull, 1952), incentive learning (Logan & Wagner, 1965), or, more generally, the representational response (Spear, 1967; Vogel, Mikulka, & Spear, 1966).

A learned r_G may affect *S*'s subsequent behavior when the goal condition is changed. Perhaps the most dramatic effect occurring under this situation is the successive contrast effect: *S*'s performance to a new reward condition is inversely related to the amount of reward previously associated with the experimental situation. That is, comparing with a control, the magnitude of *S*'s response increases (or decreases) if the new reward has a larger (or smaller) reward value. However, this effect has been found to be largely unidirectional; the switch to a small reward of given amount results in a performance decrement that is greater the larger the previous reward magnitude (Crespi, 1942; Di Lollo & Beez, 1966; Spence, 1956). This is the "depression" effect (Crespi, 1942). The "elation" effect (Crespi, 1942), as indicated by Spence, is questionable.

From the viewpoint of the sequential hypothesis (Capaldi, 1966, 1967), generalization decrement and change of stimulus intensity are assumed to be the major factors producing the depression effect. The basic assumption is that different goal conditions occasion distinctive stimuli (e.g., water, food, and nonreward produce S^W, S^R, and S^N, respectively). This distinctive stimulus may be conditioned to an instrumental response on the rewarded trials and results in a relatively permanent change in *S*'s behavior.

Generalization decrement is assumed to occur when the goal condition is changed (e.g., from reward to nonreward, large reward to small reward, or from food reward to water reward, etc.). In the changed reward situation, *S*'s performance is determined by the habit that generalizes from the previously conditioned stimulus to the newly occasioned stimulus and its stimulus intensity. Therefore, a depression effect would be predicted if the amount of generalization decrement is great and the newly occasioned stimulus intensity is weak. Moreover, since the association formed during the preshift learning situation is neither interfered with nor extinguished during the subsequent training situation, one testable implication is that if the original stimulus which conditioned to the response in acquisition was reinstated, that response would occur strongly in spite of the interpolated training.

---

[1] This report is based on a dissertation submitted to the graduate school of the University of Texas in partial fulfillment of the requirements for the PhD degree. The author wishes to express his appreciation to E. J. Capaldi for his valuable advice and assistance in the completion of this study.

[2] Requests for reprints should be sent to I-Ning Huang, who is now at Wisconsin State University at Whitewater, Department of Psychology, Whitewater, Wisconsin 53190.

The purpose of the present study was to investigate (*a*) if the depression effect would occur when the reward quality was shifted, but with reinforcing "potency" equal; and (*b*) if a particular learned r_G would be unlearned during interpolated nonreward training, small-reward-magnitude training, or training with a different quality of reward. Previous shift studies employed an experimental procedure in which the reward shift was accompanied by a decrease in the reinforcing potency of the reward, i.e., large to small reward. An attempt was made in the present study to control the reinforcing potency of the reward and to vary only the reward quality. Thus pilot work was undertaken in order to obtain a magnitude of water reward that would produce running performances comparable to that of a large food reward. A depression effect found under this condition would be of empirical as well as of theoretical interest.

METHOD

Subjects.—Sixty naive male albino rats, about 90 days old on arrival at the laboratory, were purchased from Holtzman Company, Madison, Wisconsin.

Apparatus.—The apparatus was a straight alley runway 82 in. long, 4 in. wide, and enclosed by 9-in.-high sides, covered with ½-in. hardware cloth. It had three basic sections: start, 14 in.; run, 52 in.; and goal, 16 in. It was constructed of wood and painted a mid-gray. A trial was begun by placing *S* on the 10-in. start treadle which, when depressed by *S*, started a .01-sec. clock. The front paws of *S* were always placed on the treadle's forward edge. The first clock stopped and a second started when *S* broke an infrared beam that was 4 in. from the treadle's tip (starting time). When *S* broke the second infrared beam, 52 in. from the first, the second clock stopped (running time) and a third clock started. Twelve inches from the second beam was a third infrared beam. Breaking it stopped the third and last clock (goal time) and automatically opened the sliding metal lid of a brass, 2 × 4.5 × 1.5 in. goal cup. As soon as *S* entered the goal box, a brass guillotine door 12 in. from the alley's distal end was lowered manually in order to prevent retracing. The sum of the elapsed time on the three clocks constituted the total time.

Preliminary training.—On arrival at the laboratory, *S*s were housed in individual cages. They were put on a 12-gm. Lab Blox and 20-min. watering maintenance schedule which continued throughout the experiment. The daily feeding procedures were as follows: The *S*s were first fed about two-thirds of their daily ration with the water bottles. Twenty minutes later the water bottles were withdrawn and the remaining one-third of their ration was given. On Days 1–4, all *S*s were handled individually on a wood table (2 × 3 × 4 ft.) for 5 min. before the daily feeding. From Day 5 to Day 10, *S*s were placed in an unpainted training box (16 × 4 × 9 in.), allowed to consume ten .045-gm. Noyes pellets on odd days (Days 5, 7, and 9), and allowed to drink a water bottle from the metal elbow, which was inserted through a hole (½ in. in diameter) from outside of the training box, for 40 sec. on even days (Days 6, 8, and 10).

Experimental training.—The experiment was divided into three phases. Phase 1 lasted 30 days, Phase 2, 21 days, and Phase 3, 18 days. All *S*s were run at one trial a day. At the beginning of experimental training, *S*s were randomly assigned to one of the six groups: LWL, LSL, LNL, LLL, SSL, and WWW. Group designations are given in terms of the type and magnitude of the reward (water reward, W; large food reward, L; small food reward, S; nonreward, N) received in each phase; i.e., LWL means a group received large food reward in Phases 1 and 3, with water reward in Phase 2. A trial began by removing *S* from the home cage and placing him in the apparatus in the manner described earlier. Immediately after the consumption of the reward in the goal box, either one .045-gm. Noyes pellet (S), five pellets (L), or 35-sec. water drinking (W), *S* was returned to the home cage. On a water-rewarded trial, the elbow of a water bottle was inserted in the metal goal cup from outside of the goal box. The ascending small air bubbles observed from the glass water bottle were used as an index of drinking. Drinking time was measured by a stopwatch. On a nonrewarded trial, *S* was confined in the goal box for 30 sec. All *S*s were run randomly within each group, and the six groups were run according to a prearranged random order.

Three *S*s died (two in SSL and one in LLL) and five *S*s were too weak to continue the experiment (one in SSL, two in WWW, one in LWL, and one in LSL). Two more *S*s were then run to fill in SSL and one to fill in WWW. One *S* in LNL was randomly discarded to equalize the number of *S*s in each group.

Experimental design.—The experimental design was such that (a) by comparing LWL and WWW in Phase 2, it can be seen if the depression effect occurred when the reward quality was shifted; (b) the effect of different types of interpolated training may be evaluated by comparing Groups LWL, LNL, and LSL in Phase 3 with each other as well as with the control LLL; (c) the effects of transfer shift (shift to a reward condition previously experienced) groups (LWL, LNL, and LSL) as compared to a nontransfer shift (shift to a reward condition never experienced) group (SSL) may be evaluated in Phase 3.

RESULTS

Separate time measures (in seconds) were obtained for each runway section. Since individual measures gave essentially identical statistical indications, only the running-time data are presented here.

It should be noted that the present study relies heavily on a sequential notion. The stimulus complex controlling S's behavior on the Nth trial in the runway was assumed to be composed of the outcomes that occurred on Trial $N - 1$ and the environmental stimuli present on Trial N. Therefore, Trials 2–31, Trials 32–52, and Trials 53–70 constituted the data of Phases 1, 2, and 3, respectively.

Phase 1.—Differences between the six groups on the initial trial were insignificant ($F < 1$), as were the subsequent eight daily Fs for that measure. However, an analysis of variance performed on the last 10 acquisition trials indicated that the group effect (reward condition) was significant, F $(5, 48) = 7.372$, $p < .001$. Since the trial effect was negligible, F $(9, 432) = 1.216$, $p > .20$, learning level was assumed to be asymptotic.

Further analyses employed orthogonal comparisons. The running time of the large-reward groups and the water-reward group on the last 10 acquisition trials was significantly shorter, F $(1, 48) = 35.444$, $p < .001$, than for the small-food-reward group. The large-reward groups produced running times similar to that of the water-reward group, F $(1, 48) < 1$. The mean running time was 1.2244 sec. for the water-reward group and 1.1165 sec. for large-food-reward groups (mean running times were 1.0882, 1.1283, 1.2279, and 1.0214 sec. for LSL, LLL, LNL, and LWL, respectively).

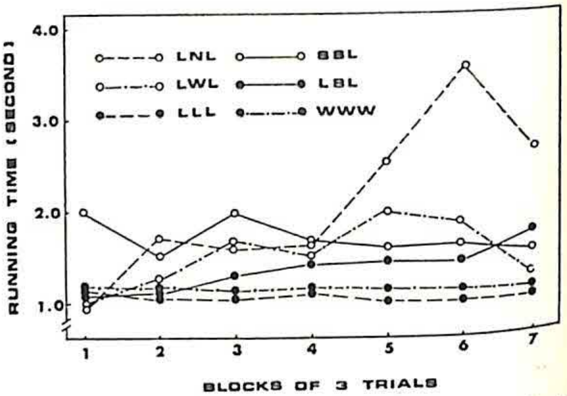


FIG. 1. Mean running time in blocks of three trials for each of the six groups on the Phase 2 training.

Phase 2.—All group means of running time for successive blocks of three trials are presented in Fig. 1.

An analysis of variance was performed between LSL and SSL for each of the seven blocks of Phase 2. There was no indication of a depression effect. LSL had a shorter running time on the first six blocks than SSL, with the first block significant, F $(1, 16) = 4.869$, $p < .05$. On Block 7, however, the running time of LSL was longer than SSL, but it was not statistically significant.

The running time of LWL was shorter than WWW on Block 1, but was longer on the subsequent six blocks. On Block 5 the difference between LWL and WWW reached statistical significance, F $(1, 16) = 5.012$, $p < .05$, and barely missed the .05 level of significance on Block 6, F $(1, 16) = 4.019$, $p = .0595$.

The delay of reward was measured for each S in LWL when the reward quality was shifted. On the first shifted trial (Trial 31), the mean delay time was 19.67 sec. ($SD = 19.82$). The delay reduced rapidly on Trial 32 ($M = 4.22$ sec., $SD = 7.08$) and was essentially near zero thereafter. Furthermore, the drinking behavior was intermittent for most of the Ss in LWL for the first few trials.

Phase 3.—Since the performance adjustments during Phase 3 for LNL, LWL, LSL,

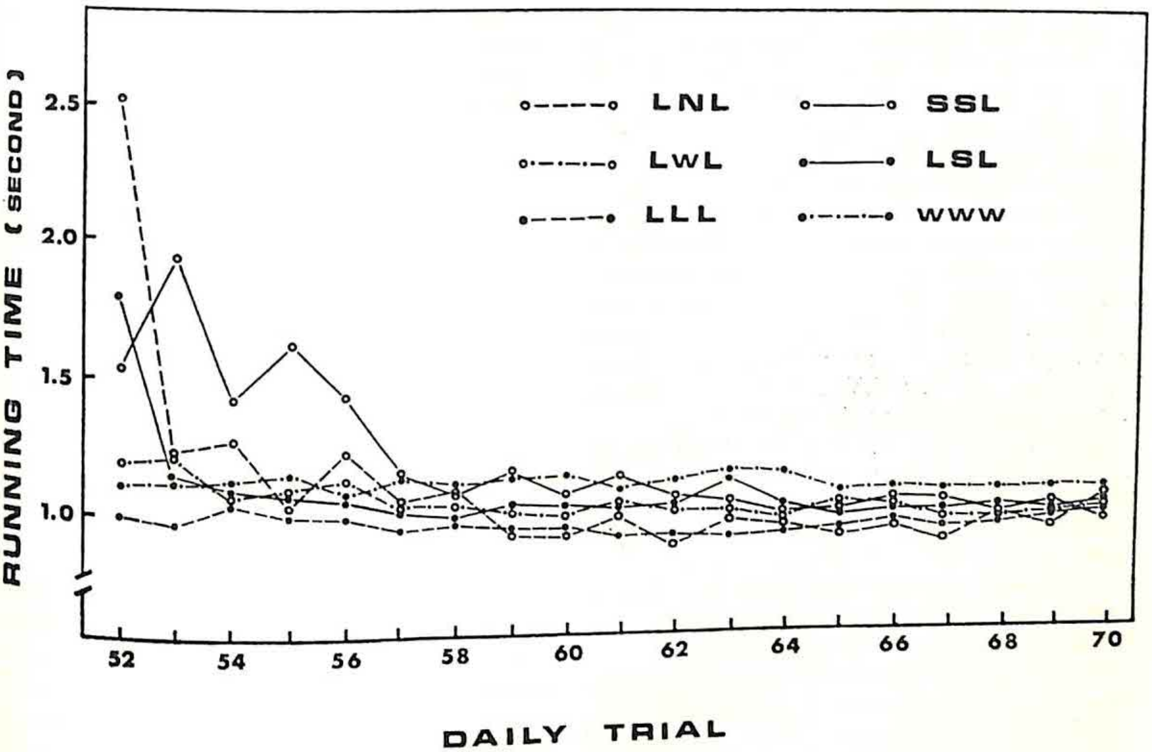


FIG. 2. Mean running time for each of the six groups on the last trial of Phase 2 (Trial 52) and the Phase 3 training.

and SSL were accomplished rather rapidly in the present experiment, the group means for each daily trial are presented in Fig. 2.

The overall Group × Trial analysis of variance indicated that the group effects were insignificant, while both Group × Trial and trial effects were highly significant beyond the .001 level.

An analysis of variance was performed among all groups except WWW for each daily trial. The running-time differences were significant on Trial 53, F (4, 40) = 3.320, $p < .05$, Trial 54, F (4, 40) = 3.291, $p < .05$, and barely missed significance on Trial 55, F (4, 40) = 2.407, $p = .0645$. Subsequent daily Fs were not statistically significant.

Further statistical analyses among these five groups employed orthogonal comparisons. A comparison of shift vs. nonshift (LSL, LNL, LWL, and SSL vs. LLL) indicated that LLL had a numerically shorter running time (all daily p's $> .05$) than the shift groups. The comparison of transfer shift vs. nontransfer shift (LSL, LNL, and LWL vs. SSL) indicated that SSL, which had been shifted to a reward never experienced, had a significantly longer running time on Trials 53, F (1, 40) = 9.502, $p < .005$; 54, F (1, 40) = 6.759, $p < .025$; and 55, F (1, 40) = 8.136, $p < .025$, than the transfer shift groups. The performance of LNL, LWL, and LSL was almost identical. Since SSL had never exceeded the performance of LLL in Phase 3, an elation effect was not observed.

DISCUSSION

A marked depression effect obtained in LWL when the reward quality was shifted during Phase 2 was quite interesting. In the present study, the large-food-reward groups and the water-reward group both produced relatively similar performance levels (running times) during Phase 1. The reinforcing potencies of water and food might be assumed, then, to be comparable. Consequently, the depression effect found in LWL might not be fully accounted for in terms of reduction of reward value

(Amsel, 1958, 1962, 1967; Spence, 1956). However, the performance decrement of LWL may be interpreted by assuming two qualitatively different and thus competitive r_Gs (r_G-food and r_G-water). Since two qualitatively different r_Gs have been associated with the common stimulus situation and only the dominant r_G during the competition would determine *S*'s behavior, at least a temporary performance decrement should be obtained. Following this analysis, a performance decrement should also be obtained in LWL when two qualitatively different r_Gs, once again, are encountered in Phase 3 training. In fact, LWL showed no sign of performance decrement during Phase 3.

Furthermore, the performance decrement of LWL in Phase 2 may not have been due to a change in the delay of reward during the first few postshifted trials. Since the performance decrement of LWL was gradual and the maximal decrement was observed late in the Phase 2 training, the delay of reward may not have direct and immediate effect on LWL.

The findings of rapid performance adjustment when the previous reward condition was reinstated, despite the use of either interpolated nonreward (Logan, 1960; Vogel et al., 1966), small reward (Capaldi & Lynch, 1967; Wagner & Thomas, 1966), or even a different reward quality training, suggest that a particular learned r_G would not be unlearned during that interpolated training phase.

The present findings support Capaldi's (1966, 1967) sequential analysis. The depression effect found in LWL during Phase 2 can be interpreted by assuming low responsivity and considerable generalization decrement when reward was shifted from food to water. It is quite reasonable to assume that the distinctive stimulus of water reward (S^W) is located further from the distinctive stimulus of large reward (S^{RL}) on the stimulus continuum (Capaldi, 1966) because water and food are two *qualitatively* different rewards. On the other hand, the stimulus intensity of water (V_W) and large food (V_{RL}) could be assumed to be roughly equal because both produced relatively similar running performances.

It may be assumed that when the reward was shifted from food to water for LWL in Phase 2, only a small part of the functional S^{RL} was replaced by S^W on each trial (low responsivity). Since the S^{RL} of the previous trials was still the major functional stimulus complex on the initial postshifted trials, generalization decrement was small and gradual and the depression effect occurred slowly. The maximal performance decrement, as well as the depression effect, would be observed on the trials (Block 5 of the present study) when S^{RL} was totally or almost totally replaced by S^W, and from then on, S^W was conditioned to the instrumental learning situation.

The LSL showed no depression effect during Phase 2. The reason might be that the difference between two reward magnitudes (one pellet vs. five pellets), as well as between the two *V*s, was so small that the amount of generalization decrement was insufficient to produce a noticeable depression effect.

The sequential hypothesis is consistent with the findings obtained in Phase 3. The hypothesis suggests that during Phase 2 training, S^{RL} was merely replaced by S^{RS} for LSL, that S^{RL} was replaced by S^N for LNL, and that S^{RL} was replaced by S^W for LWL. The association learned in Phase 1 was not directly affected during Phase 2 training. Thus, the rapid performance adjustment indicated by the three transfer shift groups, LWL, LNL, and LSL, was indeed expected during Phase 3 when the original reward condition was reinstated. It is interesting to note that the performance adjustment of the nontransfer shift group, SSL, in Phase 3 was slower than the transfer shift groups (see Fig. 2). The additional trials required by SSL to reach a performance level comparable to that of LSL, LWL, and LNL in Phase 3 was not due to slower running on the last block of Phase 2 trials. In fact, the running time of LNL or LSL was longer than that of SSL on the last block of Phase 2 trials (see Fig. 1), but LSL and LNL adjusted their performances faster than SSL in Phase 3.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

Amsel, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, 1962, **69**, 306–328.

Amsel, A. Partial reinforcement effects on vigor and persistence. In K. W. Spence & J. C. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* New York: Academic Press, 1967.

Capaldi, E. J. Partial reinforcement: A hypothesis of sequential effects. *Psychological Review*, 1966, **73**, 459–477.

Capaldi, E. J. A sequential hypothesis of instrumental learning. In K. W. Spence & J. T.

Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory.* New York: Academic Press, 1967.

Capaldi, E. J., & Lynch, D. Repeated shifts in reward magnitude: Evidence in favor of an associational and absolute (noncontextual) interpretation. *Journal of Experimental Psychology,* 1967, **75**, 226–235.

Crespi, L. P. Quantitative variations of incentive and performance in the white rat. *American Journal of Psychology,* 1942, **55**, 467–517.

Di Lollo, V., & Beez, V. Negative contrast effect as a function of magnitude of reward decrement. *Psychonomic Science,* 1966, **5**, 99–100.

Hull, C. L. *A behavior system.* New Haven: Yale University Press, 1952.

Logan, F. A. *Incentive.* New Haven: Yale University Press, 1960.

Logan, F. A., & Wagner, A. R. *Reward and punishment.* Boston: Allyn & Bacon, 1965.

Spear, N. E. Retention of reinforcer magnitude. *Psychological Review,* 1967, **74**, 216–234.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale University Press, 1956.

Vogel, J. R., Mikulka, P., & Spear, N. E. Effect of interpolated extinction and level of training on the "depression effect." *Journal of Experimental Psychology,* 1966, **72**, 51–60.

Wagner, A. R., & Thomas, E. Reward magnitude shifts: A saving effect. *Psychonomic Science,* 1966, **4**, 13–14.

(Received January 22, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 70–74

EFFECTS OF ARTICULATORY ACTIVITY AND AUDITORY, VISUAL, AND SEMANTIC SIMILARITY ON THE SHORT-TERM MEMORY OF VISUALLY PRESENTED PAIRED ASSOCIATES[1]

WILLIAM E. GUMENIK[2]
University of Toledo

Auditory similarity among stimuli significantly increased errors in short-term memory (STM) of visually presented paired associates, while visual and semantic similarity did not. Reading the pairs aloud at presentation facilitated STM, but there were no differences between conditions in which the pairs were silently read and in which tongue and lip movements were restricted. Both primacy and recency effects were found in the silent reading and restricted movement condition, but only primacy effects were found in the aloud condition. Results were interpreted as being consistent with an auditory trace rather than a motor feedback hypothesis of STM.

Several studies have shown that auditory similarity of stimuli interferes with short-term memory (STM) even when the stimuli are visually presented (e.g., Conrad, 1962, 1963, 1964, 1965, Conrad & Hull, 1964, Hintzman, 1967). Hintzman distinguished between two hypothetical sources of the effects of auditory similarity on the STM of visually presented material. One is through generalization among similar auditory traces associated with the visual stimuli. The other is through generalization along a motor rather than auditory dimension, specifically among kinesthetic feedback cues produced by similar articulatory movements in subvocal rehearsal of the visual stimuli. Hintzman found confusion errors in serial STM between letters having similar places of articulation, but low auditory confusability and interpreted his results as supporting the motor feedback hypothesis of STM.

Murray (1965b, 1966, 1967) found in several experiments that increased vocalization activity (mouthing, whispering, reading aloud) improves STM. These findings might be accounted for by an auditory trace hypothesis in terms of increased auditory cues with increased vocalization or by a motor feedback hypothesis in terms of increased vigor of articulatory movements with increased vocalization. Murray, in fact, attributed the facilitating effects of increased vocalization to both increased auditory and motor cues. However, other findings by Murray, that hearing stimuli was equally effective to voicing them in improving STM (1965c) and that white noise which interfered with *S*'s hearing his own voice decreased the advantage of vocalization on STM (1965a), could be interpreted as favoring the role of auditory rather than motor components of vocalization in STM.

The present experiment examined the effects of auditory, visual, and semantic similarity on STM of visually presented paired associates under conditions in which the pairs were read aloud at presentation, read silently, or in which movements of *S*'s lips and tongue were restricted. Both the auditory trace and motor feedback hypotheses should predict less errors in the STM of pairs read aloud than pairs read silently and more errors in the STM of pairs with acoustically identical (homonymic) than acoustically dissimilar stimulus terms. However, the motor feedback hypothesis should also predict poorer STM of pairs which were presented while tongue and lip movements were impeded, thereby interfering with articulatory movements, than pairs which were silently read with no restriction of movement. On the other hand, the auditory trace hypothesis should predict no differences between these latter two conditions.

---

[1] The author is indebted to James Brenner for his assistance in data collection and analysis and to Fay Rossman and James Lowery for their assistance in data analysis. The data were reported, in part, at the meeting of the Psychonomic Society, St. Louis, October 1968.

[2] Requests for reprints should be sent to William E. Gumenik, Department of Psychology, University of Toledo, Toledo, Ohio 43606.

METHOD

Subjects.—The *S*s were 144 introductory psychology men students randomly assigned to 12 treatment groups of 12 *S*s each.

Procedure.—Thirty-two word-number pairs were projected one at a time by a Sawyer slide projector Model 500R onto a screen 7½ ft. away. The slide projector was controlled by a Hunter Interval Cycler Model 124S. Each pair was presented for 2 sec., with .3 sec. allowed for advancement of the slide in the projector.

The pairs were presented in eight series of four pairs each. Following presentation of the fourth pair in each series, *S* was given a 3 × 5 in. card on which the four stimulus terms were printed from top to bottom with blanks along side them, in which *S* was to write the corresponding numbers. The *S* was given 30 sec. to respond and an additional 30-sec. rest period before presentation of the first pair of the next series. Within any series, the order of presentation of the pairs was randomized for each *S* and the order of the stimuli on the response card was randomly selected from two random orders which differed for each series. The order in which the series were presented was randomly determined for each *S*. A four-pair warm-up series preceded the eight test series.

Each *S* was tested under one of four stimulus similarity conditions: auditory, visual, semantic, or control (dissimilar on all three dimensions). The test series each consisted of two sets of two pairs each. The stimulus terms of the pairs within any set were either homonyms, visually similar, synonyms, or dissimilar controls, depending on similarity condition. In the case of visual similarity, the criterion used was as few letters as possible differing in their respective positions.

An attempt was made to have the four similarity conditions differ as little as possible except for the independent variable. The stimulus term of one pair of each set was the same across conditions, while the stimulus term for the other pair varied with condition. For example, if the stimulus term of the pair which was the same across conditions was PAIN, the stimulus term of the other pair in the set was PANE, PAIL, ACHE, or ARCH for the auditory similarity, visual similarity, semantic similarity, or control conditions, respectively. The stimulus terms which varied with conditions were matched on Thorndike-Lorge (1944) frequency.

The response terms were two-digit numbers, with no numbers ending in 0 or 5 or having the same digit given twice being used. The response terms were the same across similarity conditions. In order to maximize the distinctiveness of each series from the others, the same initial digit was used throughout a series, but not in any other series.

TABLE 1

MEAN ERRORS FOR FOUR CONDITIONS OF STIMULUS SIMILARITY AND THREE CONDITIONS OF ARTICULATORY ACTIVITY

Articulatory activity	Stimulus similarity				
	Homonymic	Visual	Semantic	Control	All conditions
Aloud	15.5	11.6	13.3	10.8	12.8
Silent	17.5	16.1	15.0	15.6	16.0
Restricted	17.8	14.5	14.1	16.1	15.6
All conditions	16.9	14.1	14.1	14.2	

During presentation of the pairs, *S* either read them aloud, read them silently, or extended his tongue and closed his lips and teeth on it. Irrespective of the articulatory activity condition during presentation of the pairs, during testing *S* silently wrote his responses on the response cards. Each *S* was tested individually.

The three articulatory activity conditions were combined factorially with the four stimulus similarity conditions to yield 12 between-*S* treatment conditions. Presentation position of pairs within four-pair series was a within-*S* variable.

RESULTS

Mean errors for the four stimulus similarity and articulatory activity conditions are shown in Table 1. Analysis of variance showed the stimulus similarity effect to be significant, $F\ (3, 132) = 4.01$, $p < .01$. According to Newman-Keuls tests, the homonymic condition resulted in significantly more errors at the .05 level than semantic similarity, visual similarity, or control conditions which did not significantly differ among each other. The articulatory activity effect also was significant, $F\ (2, 132) = 8.14$, $p < .01$; Newman-Keuls tests revealing that the aloud condition resulted in significantly fewer errors at the .01 level than the silent and tongue-restricted conditions which did not differ significantly among each other at the .05 level. The Similarity × Articulatory Activity interaction was not significant.

If the greater number of errors in the auditory similarity condition than the other similarity conditions was attributable to auditory generalization, this generalization should be reflected in more confusion errors between pairs in which the stimuli were homonyms than pairs in which they were not. Therefore an analysis was done on intraset confusion errors (i.e., giving the response term of one pair to the stimulus term of the other pair in the set). Insofar as the similarity conditions were concerned, the results of this analysis were similar to the analysis of total errors, the difference between the auditory similarity condition and the other similarity conditions being significant, this time at the .01 level by Newman-Keuls test. Neither the articulatory activity main effect nor the Articulatory Activity × Similarity interaction was significant at the .05 level. Also, examination of the data revealed that the proportion of total errors which were intraset confusion errors was greater for the auditory than for the other similarity conditions. These proportions were .37, .29, .27, and .25 for auditory, visual, semantic, and control similarity conditions, respectively.

Mean total errors at each presentation position within series are shown in Fig. 1 for the three articulatory activity conditions. Since there are four presentation positions, the maximum total errors possible is 8 instead of 32. The presentation position effect, F (3, 396) = 64.10, $p < .01$, and the Presentation Position × Articulatory Activity interaction, F (6, 386) = 7.88, $p > .01$, were significant; while the Presentation Position × Similarity and the triple-order interaction were not. Simple effects of presentation position at each articulatory activity condition were significant at the .01 level. Newman-Keuls tests revealed that for both the silent and restricted conditions, the first and last presentation position resulted in significantly fewer errors at the .01 level than the middle two positions. For the aloud condition, the last position resulted in significantly fewer errors at the .01 level than the remaining three positions.

Simple effects of articulatory conditions at presentation positions were significant at the .01 level at Positions 3 and 4, but were not significant at the .05 level for Positions 1 and 2. Newman-Keuls tests showed that at both Positions 3 and 4 the aloud condition

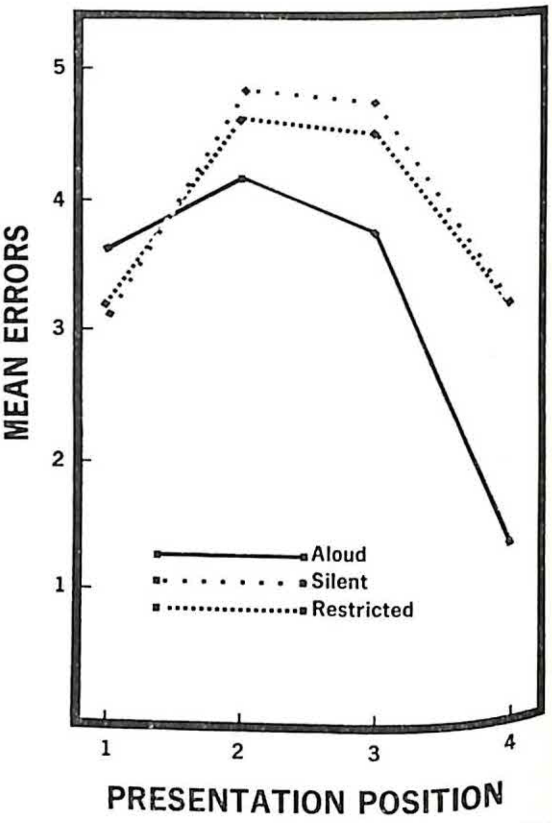


FIG. 1. Mean errors as a function of presentation position for the three articulatory activity conditions.

resulted in significantly fewer errors at the .01 level than did the silent or restricted conditions.

It was initially assumed that since S could make his responses in any order he wished, the position of the stimulus words from top to bottom on the response cards should have no effect on errors. Inspection of the data confirmed that this was the case.

DISCUSSION

The present findings of significantly more errors between visually presented pairs in which the stimuli were homonyms and therefore acoustically identical than when they were semantically similar, visually similar, or dissimilar controls are in agreement with previous studies which emphasized the contribution of acoustic similarity to errors in STM (e.g., Conrad, 1962, 1963, 1964, 1965; Conrad & Hull, 1964; Hintzman, 1967). Also, significantly fewer total errors were found when Ss read the pairs aloud at presentation than when they read them silently. This is in agreement with previous findings by Murray (1965b, 1966, 1967) on the effect of vocalization on STM. These findings are consistent with both auditory trace and motor feedback hypothesis of STM.

However, the present findings of no differences between the restricted tongue and lip movement condition and the silent reading condition, while consistent with an auditory trace hypothesis, are difficult to reconcile with a motor feedback hypothesis. It is, of course, possible that the articulatory movements involved in subvocal rehearsal of the visual stimuli are so small that extending one's tongue and closing one's lips around it would not impede them in the least. If this were to be the case, however, it would appear to be gratuitous to postulate a peripheral motor mechanism at all. Such rehearsal could be completely central. Also, if one held that all that were involved were such small movements, one might have some difficulty in accounting for the effects of vocalization in terms of increased vigor of articulatory movements. The present results do not rule out Murray's suggestion that *S* relies both on auditory and motor cues. They do indicate, however, that the postulation of such motor cues is superfluous in accounting for the present data.

The present finding of no semantic similarity effects on STM is in agreement with previous findings by Baddeley (1964, 1966; Baddeley & Dale, 1966) of no substantial semantic effects on STM and in opposition to findings by Wickens and his coworkers (Wickens & Clark, 1968; Wickens & Eckler, 1968; Wickens & Simpson, 1968) and Henley, Noyes, and Deese (1968), who did find semantic effects. A major difference between the present experiment and those of Wickens and of Henley et al. is the present use of a paired-associate technique, allowing for the separation of stimulus and response similarity. In the present experiment, semantic similarity was varied only over the stimulus terms. Perhaps the effects of semantic similarity on STM are limited to response similarity.

A study by Beecroft (1956) traced the effect of semantic similarity of stimuli on paired-associate learning from the first anticipation trial. While there are several differences between the first anticipation trial of paired-associate learning and the present technique, including differences in time between presentation and recall and the learning of material between recall of different pairs in the anticipation method, the two techniques are somewhat analogous. While Beecroft found that semantic similarity of the stimulus terms affected retention of the pairs on the first anticipation trial, this effect only occurred if four or more stimulus terms were synonymous with each other. The findings of Beecroft and the present experiment both demonstrate that the effects of semantic similarity of stimuli are so slight that having each stimulus term a synonym of one other stimulus term does not affect STM of paired associates.

No visual similarity effects on STM were found. This is in agreement with earlier findings by Baddeley (1966), using auditorily presented material, and Levy and Murdock (1968), using visually presented letters.

The findings of both primacy and recency effects, under the silent and restricted conditions, in paired-associate STM with the present technique are in disagreement with Murdock's (1963a, 1963b) findings, using a probe technique, of only recency effects in paired-associate STM. The finding of only recency effects under the aloud condition is in agreement with Murdock's findings. Further research is necessary on the determinants of the appearance of primacy effects in the STM of paired associates.

One speculative possibility stems from the fact that the probe technique allows a much shorter response period than the present technique. Shortening the time between presentation of the stimulus and *S*'s response should, of course, increase recency effects since STM drops rapidly over time and possibly obliterate any primacy effects. In the present experiment, the aloud condition, which resulted in better STM and therefore possibly faster responding, showed only a recency effect. Furthermore, the advantage of the aloud condition over the silent and restricted condition was restricted to recency effects, occurring only at the last and penultimate presentation positions.

REFERENCES

Baddeley, A. D. Semantic and acoustic similarity in short-term memory. *Nature,* 1964, **204**, 1116–1117.

Baddeley, A. D. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity, *Quarterly Journal of Experimental Psychology,* 1966, **18**, 362–365.

Baddeley, A. D., & Dale, H. C. A. The effect of semantic similarity on retroactive interference in long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 417–420.

Beecroft, R. S. Verbal learning and retention as a function of the number of competing associations. *Journal of Experimental Psychology,* 1956, **51**, 216–221.

Conrad, R. An association between memory errors and errors due to acoustic masking of speech. *Nature,* 1962, **193**, 1314–1315.

Conrad, R. Acoustic confusions and memory span for words. *Nature,* 1963, **197**, 1029–1030.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology,* 1964, **55**, 75–87.

Conrad, R. Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 161–169.

Conrad, R., & Hull, A. J. Information, acoustic confusion and memory span. *British Journal of Psychology,* 1964, **55**, 429–432.

Henley, N. M., Noyes, H. L., & Deese, J. Semantic structure in short-term memory. *Journal of Experimental Psychology,* 1968, **77**, 587–592.

Hintzman, D. Z. Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 312–316.

Levy, B. A., & Murdock, B. B., Jr. The effects of delayed auditory feedback and intralist similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 887–894.

Murdock, B. B., Jr. Short-term memory and paired associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 320–328. (a)

Murdock, B. B., Jr. Short-term retention of single paired associates. *Journal of Experimental Psychology,* 1963, **65**, 433–443. (b)

Murray, D. J. The effect of white noise upon the recall of vocalized lists. *Canadian Journal of Psychology,* 1965, **19**, 333–345. (a)

Murray, D. J. Vocalization-at-presentation and immediate recall with varying presentation rates. *Quarterly Journal of Experimental Psychology,* 1965, **17**, 47–56. (b)

Murray, D. J. Vocalization-at-presentation, auditory presentation and immediate recall. *Nature,* 1965, **207**, 1011–1012. (c)

Murray, D. J. Vocalization-at-presentation and the recall of lists varying in association value. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 488–491.

Murray, D. J. The role of speech responses in short term memory. *Canadian Journal of Psychology,* 1967, **21**, 263–275.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

Wickens, D. D., & Clark, S. Osgood dimensions as an encoding class in short-term memory. *Journal of Experimental Psychology,* 1968, **78**, 580–584.

Wickens, D. D., & Ekler, M. R. Semantic as opposed to acoustic encoding in STM. *Psychonomic Science,* 1968, **12**, 63.

Wickens, D. D., & Simpson, H. K. Semantic vs. phonetic encoding in short term memory. Paper presented at the meeting of the Midwestern Psychological Association, St. Louis, October 1968.

(Received January 23, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 75–78

EFFECTS OF EVENT PROBABILITY AND COST ON PERFORMANCE IN A CONTINUOUS MOTOR TASK[1]

ALFRED G. KLIPPLE[2] AND KING M. ROBERTS

Traffic Systems Division, United States Bureau of Public Roads, Department of Transportation

A study was made of the effects of two levels of objective event probability and two levels of objective event cost on performance in a continuous motor task. The *S*s' control of the length of a changing bar of light was affected by cost, probability, and trials. Control was not found to conform to a maximization-of-expected-value model. The *S*s tended to be less conservative than optimal when very conservative control was indicated. Also, variations in event probability were found to produce greater effects than variations in event cost.

In his review of behavioral decision theory, Edwards (1961) indicated that decision theories are traditionally concerned with static situations such as the determination of the variables affecting the elicitation of a single choice among possible alternatives. It was recognized that these models represented only first approximations of the complex and dynamic manner by which organisms make their decisions. It would seem in fact that a great deal of the human's dynamic behavior should be influenced by his estimates of the likelihood of various events occurring and by the significance which is attributed to these events.

The manner in which a human performs in a continuous motor task, e.g., should in part be influenced by his expectancies and values concerning the task in much the same way that his choices are influenced in static situations. However, the strategies that are developed in a dynamic situation have not received systematic study. It was the purpose of the present study to consider the roles of objective event probability and objective value on performance in a compensatory tracking task.

Evidence has been obtained (Miller & Meyer, 1966) which indicates that *S*s do not necessarily strictly adhere to strategies which maximize expected value. While other possible *S* strategies have been hypothesized (Coombs & Pruitt, 1960), more recently, Meyer (1967) has suggested that characteristics of the task may be responsible for deviations from maximum value strategies.

To examine this, a situation was designed where *S*s received monetary payoffs to (*a*) observe the action of a moving display, (*b*) make continuous decisions based on the significance of the display's action, and (*c*) perform motor control to compensate for the display movements. Initially, it was expected that *S*s would try to maximize the product of probability and value and would approximate the optimal control characteristics. In pilot work, however, it was noted that when a situation called for very little conservatism, *S*s tended to perform more conservatively than optimally. When very conservative control was called for, *S*s tended to be less conservative than optimal. Confirmation of this finding was sought in the present study.

The major concern here is not the overall degree of optimality at which *S*s perform. It is assumed that if the sessions were long enough or if the instructions were complete enough, *S*s would perform at an optimal level. In the present study, some understanding is sought as to how deviations from optimality can be related to the manner by which *S*s combine an incomplete knowledge of probability and cost.

---

[1] The opinions and assertions contained herein are those of the writers and are not to be construed as official or as reflecting the views of the Bureau of Public Roads. The authors acknowledge the experimental assistance of Gerald Schalestock.

[2] Requests for reprints should be sent to Alfred G. Klipple, Office of Research and Development, United States Bureau of Public Roads, Washington, D. C. 20591.

METHOD

Subjects.—The *S*s were 20 paid college students between the ages of 17 and 20. Each *S* was randomly assigned to one of four groups, with the restriction that each group was limited to 3 males and 2 females.

Apparatus.—The *S* console contained a 5-in. oscilloscope on which was displayed a horizontal bar. A positive point counter and a negative point counter were located above the console, and a bar-length control was located on the right arm of *S*'s chair. Limits consisting of two pieces of black tape on the cathode ray tube face were positioned 2¼ in. to the right and left of center along the horizontal axis.

The bar was horizontally expanded and contracted in discrete steps occurring every 2 sec. The preprogrammed sequence had 40 such changes over 80 sec. and was continuously repeated. This produced an approximately normal distribution of changes in bar size. Each change in bar size could be any of five positive or negative magnitudes. The *S* also controlled the bar length by manipulating a linear potentiometer located near his right hand.

The summed bar-length voltages drove a voltage-to-frequency converter which operated the positive point electronic counter. The rate of point accumulation displayed on the counter was a linear function of the length of the bar. With summed voltage equal to 20 v., points were accumulated at the rate of 960 (9.6¢) per minute. Each 1-v. increment or decrement changed the point accumulation by 48 points (.48¢) per minute.

When the bar expanded beyond the limits, the positive point counter stopped and a fixed count on the negative point counter was accrued. One fixed set of negative points added to the negative score each time the bar exceeded the physical limits.

The data consisted of the summed bar-length voltages. The voltage ranged from zero when the bar had no extent to 31.5 v. at the limits. This voltage was sampled by a digital voltmeter at 5-sec. intervals and printed out on a Hewlett-Packard Printer. A print-out of elapsed run time was also printed with each sample.

Procedure.—The *S*s were seated in front of the console with their eyes 36 in. away from the oscilloscope display. The *S*s were told that the bar would vary in size automatically, but that the size changes could be overridden with the manual control. The *S*s were also told that the longer the bar was within the marked limits, the faster they would accumulate positive points, but that negative points would be obtained each time the length of the bar exceeded the limits. The *S*s were instructed that the negative points would be subtracted from the positive point accumulation and that they were to control the bar so as to accumulate as many points and therefore as much money as possible.

The *E* started a masking noise source, turned the room lights off, and left the room. The *S*s could see the bar, the limits, and the positive and negative point counters.

At the conclusion of each 5-min. trial, *S* was advised of the amount of money he had made. After six trials, these amounts were totaled and *S* was paid.

The experimental design was a 2×2 factorial combining two levels of probability (high and low) and two levels of cost (high and low).

For the high cost conditions, *S* was penalized 159 points (1.59¢) each time the bar exceded the limits. The corresponding negative score for the low cost condition was 53 (.53¢). The same near normal random order of bar changes was used for each level of probability. However, the voltage changes in the high-probability condition were 4 v. greater than corresponding changes in the low-probability condition. This had the effect of making it more likely at any time that the bar would increase in size beyond the limits. For any given bar size, the rate at which positive points were accumulated could be calculated, as well as the probability that the next change in bar size would produce a negative score. Thus, the total probable payoff could be determined for every possible bar size. The four variations used here produced three voltage levels where total point accumulation was greatest. For the high-cost, high-probability condition (HC, HP) the optimal voltage to be maintained was 17.5. For the low-cost, low-probability condition (LC, LP) the optimal voltage was 25.5, and for both the high-cost, low-probability (HC, LP) and low-cost, high-probability (LC, HP) conditions the optimal voltage was 21.5. It should be noted that there was a certain flatness in the payoff distribution around the optimal value. That is, a deviation of approximately 10% from the optimal control voltage resulted in only a 6% loss in total payoff.

RESULTS

The mean of the voltage readings (reflecting average bar size) was tabulated for each condition. These are shown in Fig. 1. Also shown are the voltages for each condition which would result in the greatest overall point accumulation. An analysis of variance indicates that (*a*) mean bar sizes were shorter when costs for exceeding the limits were higher, $F\ (1, 16) = 5.23$, $p < .05$; (*b*) mean bar sizes were shorter when the probability of exceeding the limits was greater, $F\ (1, 16) = 22.11$, $p < .001$; and (*c*) mean bar sizes became shorter as the number of trials increased, $F\ (5, 80) = 3.71$, $p < .01$. None of the interactions was significant.

Each group's performance was compared by *t* tests with the optimal voltage setting for that condition. For the HC, LP condition

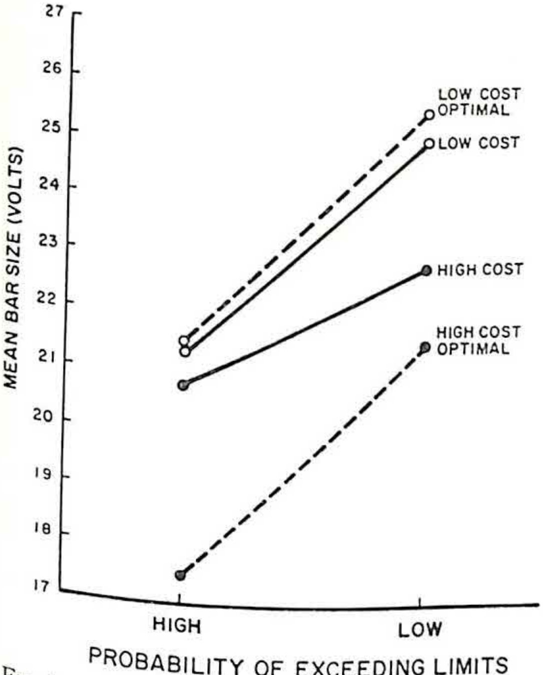


FIG. 1. Observed and optimal mean bar size as a function of probability and cost.

the mean bar size (v. $= 22.8$) was significantly larger than the optimal (v. $= 21.5$) $t(4) = 5.27, p < .05$. The mean (v. $= 20.8$) was also significantly larger than the optimal (v. $= 17.5$) for the HC, HP group, $t(4) = 9.04, p < .05$. The mean bar sizes for the LC, LP condition (v. $= 25.0$) and for the LC, HP condition (v. $= 21.4$) were slightly lower than the optimal settings (v. $= 25.5$ and 21.5, respectively). These differences were not significant.

Scheffé's test for multiple comparisons was used to determine the significance of the differences between individual groups. The LC, LP group was found to have larger bar sizes than both the LC, HP group, $F'(3, 16) = 18.3, p < .05$, and the HC, LP group, $F'(3, 16) = 24.43, p < .05$. None of the other groups differed significantly from each other.

DISCUSSION

The results show that characteristics of motor control can be related to variations in the objective probabilities and costs which are involved in the task. Increased costs for exceeding the limits resulted in more conservative control. An increase in the probability of exceeding the limits had the same effect.

As in the pilot work, *S*s were less conservative than optimal when very conservative control was called for. Also, *S*s again tended to be more conservative than optimal when the situation called for very little conservatism. However, the latter difference was not significant. It is not certain why this tendency was observed here. A likely explanation would be that since all *S*s started with essentially the same knowledge, that some mean bar size would initially be common to each group. As trials progressed and knowledge of the situation increased, *S*s should gradually change away from this mean toward an optimal setting.

The difficulty with this explanation, however, is in the significant main effect for trials and the absence of a significant Trial × Cost × Probability interaction. That is, all groups tended to become more conservative over trials. For the HC, HP group, this increased conservatism represented a change toward the optimal setting. For the LC, LP group, however, the effect represented a change away from optimal control.

Because *S*s knew nothing of the changes to expect at the start of the experiment, the optimal value described here might be somewhat unrealistic at least in the early trials. Perhaps a more appropriate explanation of strategy formation could be developed by using a continuously revised optimum as with a Bayesian model. However, the procedures of the present experiment did not permit a determination of each specific new datum and it would be difficult to apply strictly the Bayesian model here. By using a more highly controlled programming sequence, the Bayesian techniques could be used and might provide somewhat more realistic optimal values by which to compare behavior.

The individual comparisons of groups provide some information as to the relative magnitude of the effects of changes in objective cost and objective probability. In one sense the LC, LP group can be considered a control group. Optimal voltage here was 25.5. Either by making the cost high or the probability high, the optimal voltage changes to 21.5. Thus the difference in *S* control between the LC, LP condition and the LC, HP and HC, LP conditions would be about the same. Here, however, a significant difference was found between the LC, LP and the LC, HP conditions, but not between the LC, LP and the HC, LP conditions. This indicates that in this task, probability variations produce greater effects on *S* control than do similar variations in cost.

In the same sense, the HC, HP condition can be considered as a control group. In this case the optimal voltage was 17.5. Changing either the cost or probability to low increases the optimal voltage to 21.5. Neither of these changes produced significant results, although again a greater difference resulted from the variation in probability.

As in static decision making, characteristics of continuous motor control can be related to the costs and probabilities of desirable and undesirable events. However, the strategies which were developed here do not strictly conform to a maximization-of-expected-value model. It will be necessary to collect additional data to specify more accurately the strategies which are developed in this type of task as well as the particular relevant independent variables.

REFERENCES

COOMBS, C. H., & PRUITT, D. G. Components of risk in decision making: Probability and variance of preferences. *Journal of Experimental Psychology*, 1960, **60**, 265–277.

EDWARDS, W. Behavioral decision theory. In P. R. Farnsworth, O. McNemar, & Q. McNemar (Eds.), *Annual review of psychology*. Palo Alto: Annual Reviews, 1961.

MEYER, D. E. Differential effects of knowledge of results and monetary reward upon optimal choice behavior under risk. *Journal of Experimental Psychology*, 1967, **75**, 520–524.

MILLER, L., & MEYER, D. E. Decision effectiveness as a function of the structure and number of alternatives at a choice point. *Proceedings of the 74th Annual Convention of the American Psychological Association*, 1966, **1**, 25–26.

(Received January 27, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 79–87

EFFECTS OF REWARD INCREASE AND REDUCTION IN THE DOUBLE RUNWAY[1]

STUART A. KARABENICK[2]

Eastern Michigan University

Eighty-two rats received 60 trials in a double runway. On Trials 1–30, reward magnitudes were either consistently large (L) or small (S) in the first goal box (GB_1) and L or S in the second goal box (GB_2), using a 2×2 factorial design. On Trials 31–60, half of the *S*s in each preshift group were shifted to the opposite magnitude in GB_1, the GB_2 magnitude remaining at the preshift level for all *S*s. Second runway (R_2) performance rate increased for *S*s shifted from L to S magnitudes in GB_1, but the rate was not significantly above that of *S*s constantly receiving the small GB_1 reward. The *S*s shifted from S to L magnitudes in GB_1 decreased their R_2 response rate. Preshift R_2 performance and shift effects varied as a function of GB_2 magnitude. The results are discussed in terms of Amsel's frustration theory and two theories involving the concept of persevering response tendencies.

Several studies have shown that reducing the amount of reward in the first goal box (GB_1) of a double runway to zero results in an increase in second alley (A_2) response speed (e.g., Amsel & Roussel, 1952; Peckham & Amsel, 1964). In addition, *S*s shifted to zero reward in GB_1 have been shown to respond faster in A_2 than control groups that were never rewarded in GB_1 (termed a "frustration effect"; McHose, 1963; Wagner, 1959). However, reward reductions to nonzero magnitudes have failed to demonstrate shift effects comparable to those with reductions to zero reward (Barrett, Peyser, & McHose, 1965; Daly, 1968). Control groups function to provide base lines for determining the effects on A_2 speed of shifts in GB_1 magnitude as opposed to A_2 speed differences due to the absolute magnitude of reward consumed in GB_1 (Daly, 1968; McHose, 1963). One common feature among studies of reward reduction has been the use of a *single* reward magnitude in the second goal box (GB_2), typically equal to or smaller than the reward used in GB_1. This parametric constancy would appear to limit the generality of double-runway shift effects. One purpose of the present study was to extend previous studies of reward reduction by using more than one magnitude of reward in GB_2.

A second aspect of the present study was to determine whether an increase in GB_1 reward magnitude would yield shift effects in A_2 when compared to control groups consistently receiving the large reward magnitudes. It might be hypothesized that the shift effect in the case of reward increase would lead to slower A_2 responding by the shift group compared to the control group due to the reward magnitude contrast experienced by the shifted *S*s (Pubols, 1960). In the only study of the obverse of reward reduction, Bower (1962) found that an increase in the amount of reward received in the first goal compartment of a triple runway resulted in slower responding in succeeding runway sections. However, the absence of proper control groups that had always received the larger reward magnitudes precluded an analysis of any shift effect.

In order to assess the effect of an increase and decrease in the amount of reward in GB_1 and the effect of different magnitudes of reward in GB_2 on A_2 response strength, a factorial design was used in which the factors consisted of GB_1 and GB_2 reward magnitudes. During the training or "preshift" phase, *S*s in a given group received one of four reward magnitude combinations: either a large (L) or a small (S) magnitude of reward in GB_1 and either an L or an S reward in GB_2. In the second, or "postshift" phase, half of the *S*s in each group were switched to the opposite magnitude (S to L or L to S) in GB_1, while the other half continued to receive the preshift amount. This design is similar to that used by McHose (1963) and Daly (1968) in that the shifted groups continued to receive the reduced or increased reward over all postshift trials.

---

[1] This study was based on a doctoral dissertation submitted to the Rackham Graduate School of the University of Michigan. The research was supported by a predoctoral National Institute of Mental Health fellowship. The author wishes to thank David Birch, John W. Atkinson, and Edward L. Walker for their invaluable contributions and assistance.

[2] Requests for reprints should be sent to Stuart A. Karabenick, Department of Psychology, Eastern Michigan University, Ypsilanti, Michigan 48197.

METHOD

Subjects.—The *S*s consisted of 88 male hooded rats obtained from the Maxfield Supply Company (Cincinnati, Ohio) at approximately 75 days of age. Six *S*s failed to complete all trials. Their data were eliminated from all analyses.

Apparatus.—The apparatus consisted of two straight alley runways in series, where the goal compartment of the first also served as the starting compartment of the second. The entire double runway was constructed of wood floor and sides with a hinged, translucent $\frac{1}{8}$-in. Plexiglas top. Start and retrace doors were also made from translucent Plexiglas that was left unpainted. The inside dimensions of the runway were 4 in. wide and 6 in. high throughout. The lengths of the various runway components were as follows: A_1 starting compartment = 10 in.; A_1 = 45 in.; GB_1 (A_2 starting compartment) = 15 in.; A_2 = 45 in.; GB_2 = 17 in. The food cups, $\frac{1}{2}$ teaspoons mounted on the sides of both goal compartments, were located 14 in. beyond the retrace doors of each goal compartment, 1 in. above the floor and extended 1 in. into the runway. A_1 and A_2 doors were motorized, while the retrace doors were operated manually. All doors were hinged at the top so that they operated completely within the closed runway.

The runway floor was painted medium gray throughout, as were the sides of the A_1 starting compartment. The sides of A_1 and GB_1 were lined with white plastic side panels; A_2 and GB_2 were lined with black side panels. Timing circuits were controlled by photoresistive cells mounted beneath the runway floor. Each was covered with a $1\frac{1}{4}$-in. plastic disc painted the same color as the runway floor and scored to admit sufficient light with which to produce resistance changes when the shadow of *S* passed over it. Using these cells eliminated the need for any concentrated light beam, the necessary illumination coming from the fluorescent room lights diffused by the runway top. All control and timing apparatus was located in a room separate from that which contained the runway.

Deprivation schedule.—Upon receipt, *S*s were placed in individual cages and given ad-lib food and water for 26 days. This was followed by 25 days of restricted feeding at 15 gm. of food per day. After 10 days, the ration of each animal was adjusted to bring *S* to 80% ad-lib weight as determined by each *S*'s weight during the last week of ad-lib feeding. Prior to adaptation trials, *S*s were divided into two groups equated for weight and placed on a schedule that alternated restricted feeding days with total deprivation. During testing, *S*s were fed their adjusted ration after performing on runway trial days. Any uneaten ration was removed at the same time from *S*s to be run on the following day. This procedure produced a combination weight control and 22-hr. time of deprivation schedule.

Handling and runway adaptation.—The *S*s were gentled daily from the time of receipt to the beginning of runway adaptation. They were given additional handling for 5 min. after each adaptation trial. Six runway adaptation trials, one trial per day, were given with all equipment operating, but without food in place. After each trial, they were removed, weighed, and placed on a play table with a supply of 45-mg. Noyes food pellets. After 7 min., *S*s were returned to their home cages and fed approximately 30 min. later. On the first two trials, *S*s were allowed to explore each runway individually for 5 min. On Trial 3, *S*s were placed in the starting compartment of A_1, the start door was opened 5 sec. later, and *S* was given 1 min. to explore A_1. If at the end of that period *S* was in GB_1, the retrace door was lowered, the A_2 start door was opened 5 sec. later, and *S* was allowed 1 min. to explore A_2. If *S* was not in GB_1, he was removed from A_1 and placed there.

Trials 4, 5, and 6 were similar except that the retrace door was closed any time *S* entered GB_1. On the last three adaptation trials, the retrace door was also closed if *S* entered GB_2 at any time during the 1-min. exploration of A_2. If after 1 min. of exploration *S* had failed to enter GB_2, he was placed there for 10 sec. before being removed from the runway.

Experimental design.—There were four groups during the preshift phase. Each received either 1 or 15 45-mg. Noyes food pellets in GB_1 and either 1 or 15 pellets in GB_2. The reward amount for each group remained constant over the 30 preshift trials. Designating 1 pellet as a relatively small (S) reward and 15 pellets as a relatively large (L) reward, the four preshift groups can be represented, along with the final *N*s, as S:S ($N = 19$), S:L ($N = 20$), and L:S ($N = 22$), L:L ($N = 21$), where the first initial represents the relative amount of reward in the first goal compartment and the second initial represents the relative amount of reward in the second goal compartment.

After 30 trials, *S*s in each preshift condition were divided into two groups matched according to their performance in A_2 over the last 10 preshift trials. Half of the *S*s in each group were shifted to the opposite magnitude in the first goal compartment, resulting in four shifted and four nonshifted control groups. The control groups are designated S-S:S ($N=9$), S-S:L ($N=10$), L-L:S ($N=11$), and L-L:L ($N=11$). The shifted groups are designated S-L:S ($N=10$), S-L:L ($N=10$), L-S:S ($N=11$), and L-S:L ($N=10$). The *S*s were given 30 postshift trials.

Assignment to groups and trial procedure.—The *S*s were first classified into 22 homogeneous groups of four *S*s each on the basis of their 100-day-old ad-lib weight. Within each of these groups, *S*s were then assigned to one of the four preshift reward conditions. The total number of *S*s in each reward condition was then divided into two groups that alternated trial and feeding days. The *S*s were run in two squads on each day. They were given one trial every other day.

A trial began by placing *S* in the A_1 starting compartment facing the door. The start door was opened 5 sec. later without regard to *S*'s orientation. Starting latency in A_1 (S_1) was measured from the opening of the start door until *S* passed over the first photocell 6 in. from the start door. The retrace door was lowered to prevent retracing, and running time in A_1 (R_1) was measured from the time of crossing the first photocell until *S* crossed the second photocell 48 in. farther down A_1 and 5 in. in front of the food cup in GB_1. The goal box retrace door was closed to prevent retracing after *S* passed under it.

The A_2 start door was opened 5 sec. after *S* had finished the first reward, again without regard to *S*'s orientation to the start door. The starting latency in A_2 (S_2) was taken from the opening of the A_2 start door until *S* crossed over a photocell 6 in. into A_2. The timing of running time in A_2 (R_2) and the retrace door procedure were the same as in A_1.

The *S*s were removed from GB_2 5 sec. after completing the reward. They were then weighed and returned to their home cages. The *S*s were fed after an entire squad had been run, which resulted in an average of 45 min. between runway trials and feeding for *S*s within a squad.

RESULTS

All of the data were analyzed using $\log_{10} + 1.0$ conversions of the obtained time scores. Variance analyses with unequal *N*s used the unweighted means solution (Winer, 1962, pp. 375–377).

Preshift Performance

Alley 1.—Mean log starting latencies and log running times in A_1 for the four preshift reward conditions are presented in Fig. 1. To analyze the effect of first and second goal compartment magnitudes on A_1 performance, $2 \times 2 \times 10$ variance analyses (GB_1 Magnitude $\times$ GB_2 Magnitude $\times$ Trials) were performed using the last 10 preshift trials (21–30; Trial Blocks 5 and 6). The main effect for GB_1 magnitude was highly significant on both A_1 performance measures: S_1, $F\ (1, 78) = 16.74$, $p < .001$; R_1, $F\ (1, 78) = 8.26$, $p < .01$. Although the main effect for GB_2 reward magnitude was not significant: S_1, $F\ (1, 78) = 2.97$; R_1, $F\ (1, 78) = 2.44$; there was a significant interaction effect between GB_1 and GB_2 magnitude on both A_1 performance measures: S_1, $F\ (1, 78) = 10.09$, $p < .01$; R_1, $F\ (1, 78) = 5.57$, $p < .05$. As can be seen in Fig. 1, the reward in GB_2 affected A_1 performance when the GB_1 magnitude was small, but not when it was large. A Newman-Keuls test with $\alpha = .05$ substantiated that Group S:S performed significantly slower than all other groups, but the performance of the other groups did not differ significantly over the last 10 preshift trials.

Tests using S_1 indicate that all groups were improving over Trials 21–30, $F\ (9, 702) = 3.52$, $p < .01$. However, R_1 shows no such significant change, nor were there any significant interactions on either measure with trials. This indicates that the relative performance of the four conditions was consistent over the last 10 preshift trials.

Alley 2.—Mean log starting latencies and running times in A_2 are also presented in Fig. 1. The $2 \times 2 \times 10$ variance analyses (GB_1 Magnitude $\times$ GB_2 Magnitude $\times$ Trials) over Trials 21–30 were also used to evaluate A_2 performance data. That A_2 performance was an inverse function of the amount of food consumed in the first goal compartment is substantiated by the highly significant GB_1 magnitude main effect: S_2, $F\ (1, 78) = 34.76$, $p < .001$; R_2, $F\ (1, 78) = 16.78$, $p < .001$. A_2 performance was also a function of the amount of reward in GB_2, as evidenced by the significant GB_2 magnitude main effect: S_2, $F\ (1, 78) = 25.77$, $p < .001$; R_2, $F\ (1, 78) = 17.75$, $p < .001$.

The interaction effect between the GB_1 and GB_2 magnitudes was not significant using S_2, $F\ (1, 78) = 2.34$. Paired comparison of the four conditions (Newman-Keuls, $\alpha = .05$) indicated that *S*s in Group S:L performed faster in A_2 than did *S*s in all other conditions. In addition, Group L:S performed slower than all other groups. The performance of Groups S:S and L:L did not differ significantly from each other. The interaction between GB_1 and GB_2 food magnitudes was significant for the R_2 measure, $F\ (1, 78) = 4.96$, $p < .05$. Paired comparisons of the four groups indicated again that Group S:L performed significantly faster than all other groups, but the performance of all other conditions did not differ.

As with A_1, A_2 starting latency showed a significant change over the last 10 preshift trials, $F\ (9, 702) = 3.98$, $p < .01$, but the running measure did not, $F\ (9, 702) = 1.49$. There were no significant interactions involving trials for either S_2 or R_2, again an indication that the four preshift reward groups remained stable relative to each other over Trials 21–30.

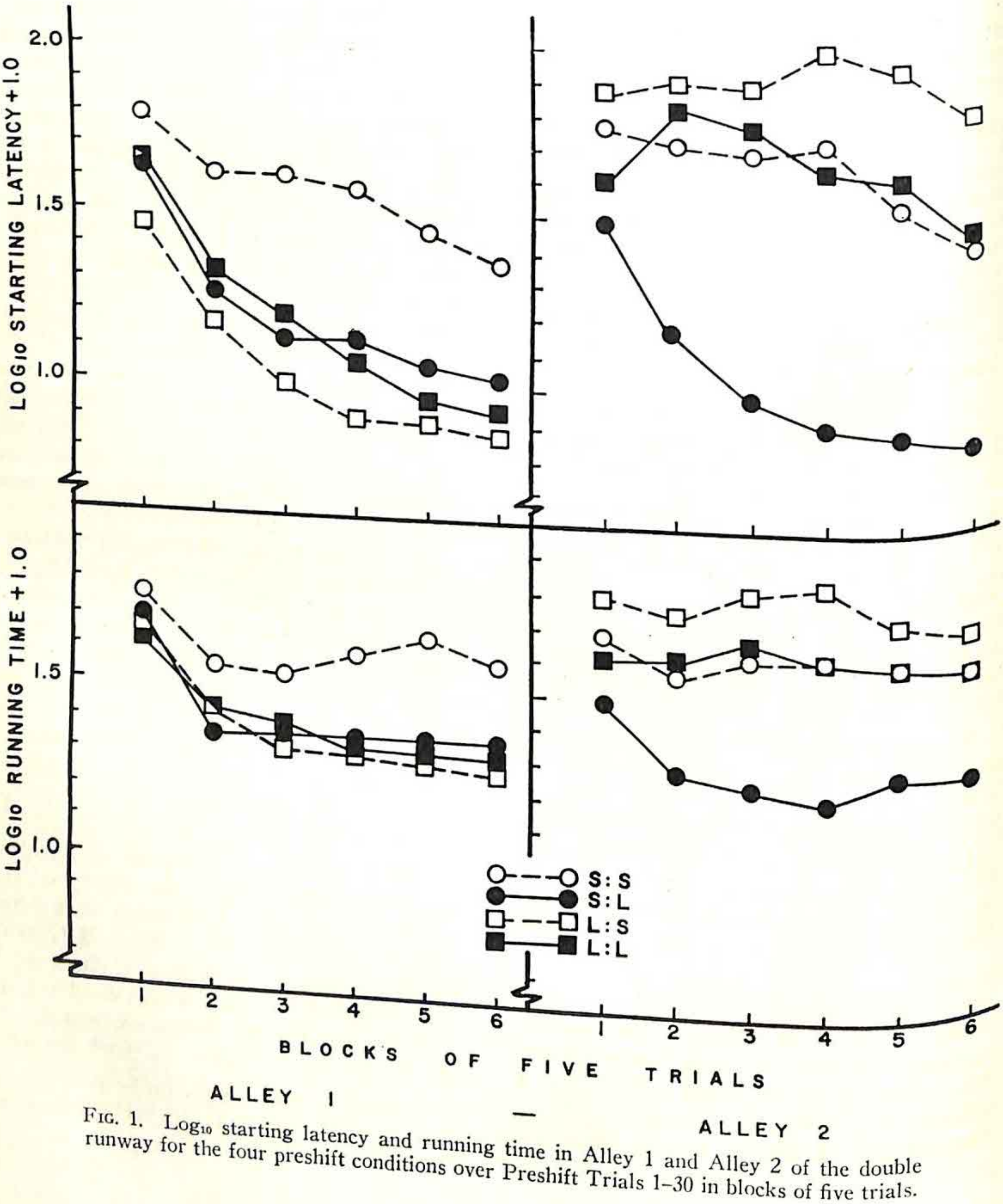


FIG. 1. Log_{10} starting latency and running time in Alley 1 and Alley 2 of the double runway for the four preshift conditions over Preshift Trials 1–30 in blocks of five trials.

Summary of preshift performance.—The rate of performance in A_1 was found to be a direct function of the magnitude of reward in GB_1. The amount of reward in GB_2 had an effect on A_1 performance when the first reward was small, but not when it was relatively large. Preshift performance rate in A_2 was an inverse function of the magnitude of reward consumed in GB_1 and a direct function of the amount of reward received in GB_2. In addition, while there was evidence for an interaction effect between GB_1 and GB_2 magnitude on A_2 running time, the interaction effect was not present on the A_2 starting measure.

Success of Group Matching

The success with which the nonshifted control and the shifted groups had been equated for preshift performance over Trials 21–30 was tested by two-way variance analyses (Shifted Subgroups vs. Nonshifted Control Subgroups × Trials) on both A_1 and A_2 performance. Matching appears to have been adequate as there were no significant differences between shifted and nonshifted subgroups. The only significant effect was an interaction for Group L:S on S_1, $F\ (9, 180) = 2.51$, $p < .05$.

Postshift Performance

Alley 1.—The starting latency and running time performance in A_1 for the four control and four shifted groups is presented in Fig. 2 for the last five preshift trials and all postshift trials (31–60) in blocks of five trials.

There were no significant differences over any postshift trial segment between the shifted and control groups that received a large reward magnitude in GB_2 (S-S:L vs. S-L:L and L-S:L vs. L-L:L). However, groups receiving a small A_2 magnitude were affected by the shift. Group L-S:S slowed relative to Group L-L:S and began to approach the level of Group S-S:S over the first 10 postshift trials. The convergence of Groups L-S:S and S-S:S is indicated by a significant Groups × Trials (31–40) interaction using R_1, $F\ (9, 162) = 3.15$, $p < .01$, although the interaction using S_1 was not significant, $F\ (9, 162) = 1.61$, $p < .10$. The performance rate of Group S-L:S increased compared to that of Group S-S:S over Trials 31–40 and began to approach the level of Group L-L:S. This approach is indicated by a significant Groups × Trials (31–40) interaction using S_1, $F\ (9, 171) = 3.68$, $p < .01$, although the effect was not significant using the R_1 measure, $F\ (9, 171) = 1.53$, $p > .10$. Subsequent to the first 10 postshift trials, there were no significant differences between the shifted groups (L-S:S and S-L:S) and their respective control groups (S-S:S and L-L:S).

Alley 2.—Mean log A_2 starting latencies and A_2 running times for the last preshift and all postshift trials were also presented in Fig. 2. A_2 performance of the eight postshift groups was analyzed in the following ways: (*a*) Performance changes from the last preshift (30) to the first postshift (31) trial were compared for shifted and nonshifted subgroups of each preshift condition using two-way variance analyses (Groups × Trials 30 vs. 31); (*b*) the performance of shifted and nonshifted subgroups was compared over Trials 31–40 and over all postshift trials (31–60); (*c*) comparisons were made over Trials 31–40 and 31–60 between the performance of groups that received a shift in GB_1 magnitude and the appropriate nonshifted control groups that had consistently received a small or large GB_1 reward.

A decrease in GB_1 magnitude produced an immediate increase in the rate of A_2 performance by Group L-S:S compared to Group L-L:S from Trial 30 to 31 as indicated by a significant Groups × Trials interaction on both A_2 performance measures: S_2, $F\ (1, 20) = 6.37$, $p < .05$; R_2, $F\ (1, 20) = 13.55$, $p < .01$. Group L-S:S continued to perform faster than L-L:S over Trials 31–40: S_2, $F\ (1, 20) = 10.16$, $p < .01$; R_2, $F\ (1, 20) = 8.29$, $p < .01$; and over all postshift trials: S_2, $F\ (1, 20) = 5.67$, $p < .05$; R_2, $F\ (1, 20) = 5.36$, $p < .05$.

There was some indication that Group L-S:S performed faster than the control Group S-S:S on the first shift trial, but the difference was not significant: S_2, $F\ (1, 18) = 3.69$, $p < .10$; R_2, $F\ (1, 18) = 3.44$, $p < .10$.[3] Although there is an apparent trend shown in Fig. 2, Shifted Group L-S:S did not perform significantly faster than Control Group S-S:S over Trials 31–40: S_2, F $(1, 18) < 1.00$; R_2, F $(1, 18) = 2.05$, $p > .10$; or over all postshift trials: S_2, F $(1, 18) < 1.00$; R_2, F $(1, 18) = 1.08$, $p > .10$.

---

[3] Single-trial simple-effects comparisons used the error term based on the first 10 postshift trials (see Winer, 1962, p. 310).

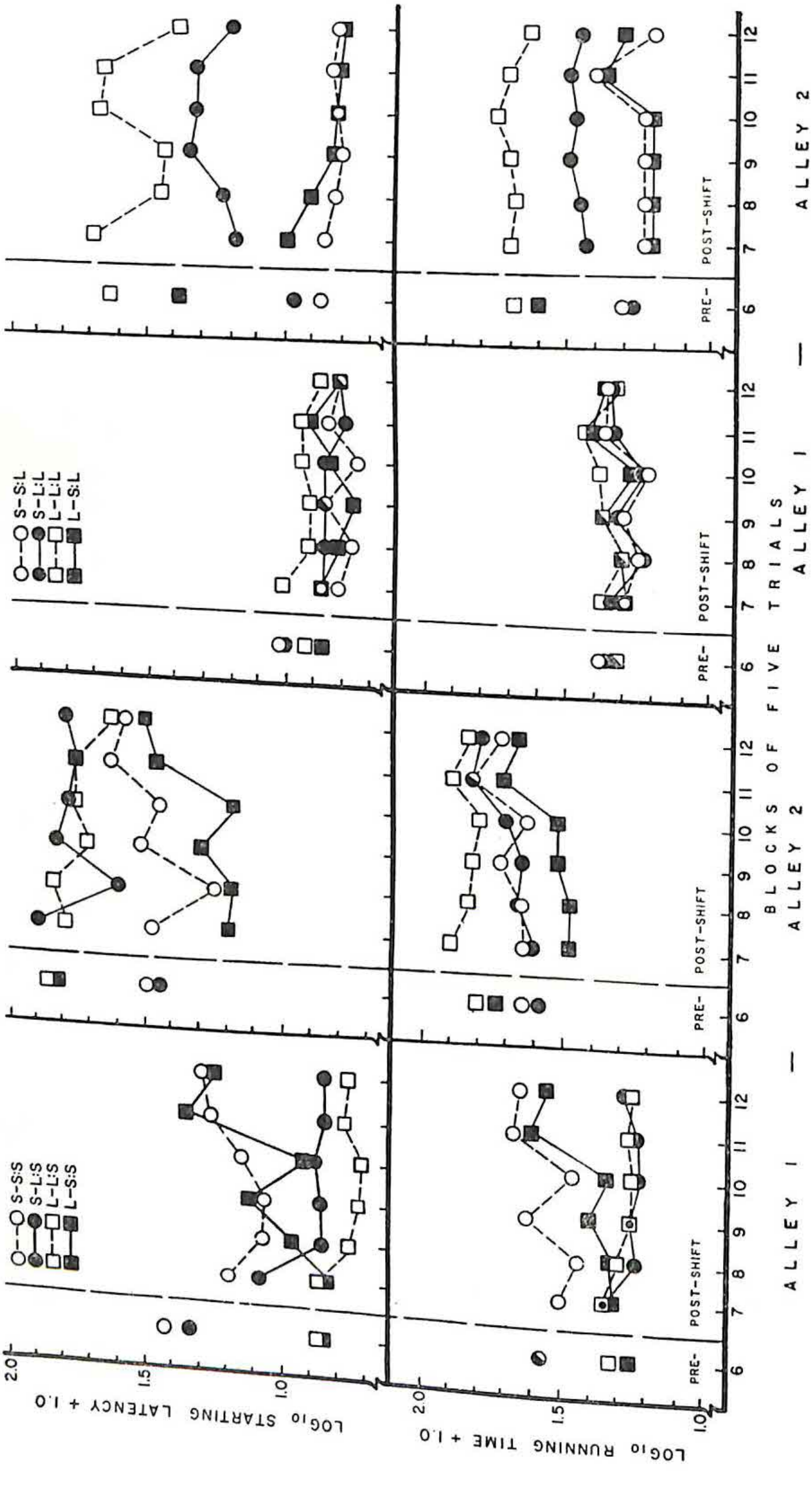


Fig. 2. Log_{10} starting latency and running time in Alley 1 and Alley 2 for the four shifted and four control groups on the last block of five preshift trials (26–30) and 30 postshift trials (31–60) in blocks of five trials. (The two left-hand panels present data for groups with a small reward in GB_2. The two right-hand panels are for groups with a large reward in GB_2.)

In the case of a decrease in GB_1 magnitude where the GB_2 magnitude was large, performance for Group L-S:L increased compared to that of Group L-L:L from Trial 30 to 31 on the running measure as shown by a significant Groups × Trials interaction, $F (1, 19) = 30.09$, $p < .01$, but not on the starting measure. Group L-S:L continued to perform faster than Group L-L:L over all postshift trials: S_2, $F (1, 19) = 14.48$, $p < .01$; R_2, $F (1, 19) = 11.46$, $p < .01$. There was, however, no indication that Shifted Group L-S:L performed faster than Nonshifted Control Group S-S:L. In fact, there was evidence to indicate that Group L-S:L was slower to leave the A_2 starting compartment than Group S-S:L over the first 10 postshift trials, $F (1, 18) = 4.59$, $p < .05$. The performance of the two groups was the same for the remainder of the postshift trials.

An increase in GB_1 magnitude resulted in a decrease in the rate of performance by Shifted Group S-L:S compared to Nonshifted Group S-S:S on the first postshift trial, but the Groups × Trials interaction was not significant for either A_2 performance measure. As seen in Fig. 2, Group S-L:S started more slowly than Group S-S:S over Trials 31–40: S_2, $F (1, 17) = 4.92$, $p < .05$; but there were no significant differences between these groups for the running measure over the series of postshift trials. There were no significant differences between Shifted Group S-L:S and Nonshifted Control Group L-L:S evaluated over all postshift trials using either A_2 performance measure.

There was also a decrease in A_2 performance for Shifted Group S-L:L relative to Group S-S:L on the first postshift trial, but the Groups × Trials interaction was not significant using either A_2 performance measure. The two groups were significantly different, however, over all postshift trials: S_2, $F (1, 18) = 9.42$, $p < .01$; R_2, $F (1, 18) = 6.23$, $p < .05$. Despite the decrease in A_2 performance, there was a tendency for Group S-L:L to continue to perform faster than Nonshifted Control Group L-L:L. The difference between these groups over all postshift trials fell just short of being significant on the R_2 measure, $F (1, 19) = 4.07$, $p < .10$.

Summary of Postshift Performance

In general, a decrease in GB_1 food magnitude resulted in an increase in the rate of A_2 performance. However, the increase in performance by the shifted groups did not exceed the level of the nonshifted control groups. An increase in GB_2 food magnitude resulted in a decrease in A_2 performance. However, there was a tendency for the shifted groups to continue to perform faster than their nonshifted control groups throughout the series of postshift trials.

DISCUSSION

A reduction in GB_1 reward magnitude resulted in significantly faster A_2 performance; however, there was no evidence for a "frustration effect" in that both groups experiencing the reward decrease failed to perform significantly faster than control groups that had consistently received the reduced GB_1 magnitude during the series of preshift and postshift trials. This result is consistent with previous studies of nonzero reward reduction that used appropriate control groups (Barrett et al., 1965; Daly, 1968).

One explanation of the failure to find evidence of a shift effect (shift group performing faster than its nonshift control group) for reductions to nonzero reward magnitudes in the present study has been suggested by Daly (1968). The argument involves the conceptualization of the double runway as operationally similar to differential conditioning in which *S*s are exposed to more than one magnitude of reward. Following Amsel's frustration theory, it is proposed that *S*s receiving different magnitudes of reward in GB_1 than in GB_2 would experience frustration on encountering the smaller of the two magnitudes. Assuming this to be the case in the present study, frustration would occur in GB_1 for Group S:L on preshift trials and for Control Group S-S:L on postshift trials. Group S-S:L would therefore perform faster in A_2 due to the presence of the additional frustration-produced drive component than would be the case if GB_1 and GB_2 reward magnitude effects were independent of each other. Since both the shift and control groups would experience frustration in GB_1, they would be responding under approximately equivalent levels of drive in A_2, and their A_2 performance levels would not be expected to differ on postshift trials.

Evidence for the presence of frustration-produced drive in A_2 for Group S-S:L could be inferred from the high response rate of Group S:L in A_2 during the series of preshift trials. A_1 performance would appear to provide contrary evidence. If frustration did occur in GB_1, anticipatory frustration (r_F) should have become conditioned to the cues of A_1. The inhibitory effect of r_F should thus have acted to slow the response rate of Groups S:L and S-S:L in A_1. This did not occur as Group S:L responded as fast as Groups L:S and L:L. However, the absence of a frustration-produced inhibitory effect for Group S:L cannot be totally ruled out since the facilitation effect of the large reward in GB_2 that was found to generalize to A_1 may have cancelled out the inhibitory effect. This interpretation was used by Daly (1968) to explain the absence of an inhibitory effect under similar reward conditions.

The conditions of reward-contrast-produced frustration would not exist for Control Group S-S:S as proposed for Control Group S-S:L. However, Group L:S would be assumed to experience frustration on encountering the smaller reward in GB_2 during preshift trials. Since r_F would become conditioned to the cues of A_2, its presence would be expected to depress the A_2 response rate for this group. Accordingly, Shifted Group L-S:S would be assumed to be responding in the presence of inhibition not present for Control Group S-S:S. Inhibition in A_2 might then have partially offset the response strength increment in Group L-S:S due to frustration-produced drive in GB_1 cancelling or partially cancelling the shift effect.

The preshift data provide some evidence inconsistent with this analysis. Given that frustration was present in GB_2 for Group L:S, it would follow that an inhibitory effect of r_F developed on preshift trials would be greater in A_2 (i.e., closer to GB_2 where the frustration occurred) than in the A_2 starting compartment (GB_1). Instead, just the opposite would appear to have been the case since Group L:S took significantly longer to leave GB_1 than all other groups, but did not run significantly slower than all other groups once in A_2. It is possible that the difference between the two measures is due to greater variance among *S*s on the running measure. However, this seems unlikely since the mean square error for R_2 (.6353) was less than half as large as that for S_2 (1.6603) for the analysis over Trials 21–30.

In addition to frustration theory, there are two related theoretical approaches that can be brought to bear to account for shift effects in the double runway. Both of these involve the notion of perseverative effects carrying over from A_1 to A_2. According to one position, Bower (1962) postulated that the strength of the response tendency in A_2 varies as an increasing function of the strength of the fractional anticipatory goal response (r_G) elicited by A_1 runway cues and a decreasing function of the amount of consummatory responding (R_G) taking place in GB_1. Atkinson (1964) and Birch (1968) have proposed a similar theory in which the primary concept is "inertial motivation." Like the persistence of r_G, any response tendency (T) built up through exposure to instigating cues in A_1 and not reduced by consummatory behavior in GB_1 would be assumed to persist and to increase the final strength of T in A_2. A_1 performance is crucial for these theories because it provides information as to the relative strength of r_G or T in A_1 prior to their reduction in GB_1.

Since both theories involve the concept of persistence, they account for the reward magnitude shift effects in similar ways. First, the failure to find a shift effect for Group L-S:L would be assumed to result from the high response rate of Control Group S-S:L, with which it was compared. However, instead of invoking frustration, the high A_1 rate would be seen as due to the continual instigation of r_G or T in A_1 and its minimal reduction (and perhaps its continual instigation) in GB_1. This interpretation is consistent with the preshift A_1 data where a large GB_2 reward magnitude was shown to affect the A_1 response rate to the extent that Group S:L performed at the same rate as Groups L:S and L:L in A_1. Thus, although the strength of r_G or T would not be reduced for Shifted Group L-S:L on postshift as on preshift trials leaving a greater resultant response strength in A_2 on postshift trials, no shift effect would be expected because of the high level of r_G or T and thus the high rate of Control Group S-S:L.

However, both perseverative theories would have difficulty accounting for the failure to find a shift effect for Group L-S:S. In both analyses, r_G or T would increase due to instigation in A_1 because of the large reward in GB_1 for Group L:S on preshift trials. A reduction in GB_1 magnitude for Shifted Group L-S:S would then be expected to leave a greater level of unreduced r_G or T to persist into A_2 than would be the case for Control Group S-S:S for which instigation in A_1 had been minimal (cf. preshift A_1 performance for Group S:S).

Accordingly, Group L-S: S should have performed faster in A_2 on postshift trials than Group S-S:S, which was not the case.

The effects of increasing the magnitude of reward in GB_1 were generally to reduce the rate of performance in A_2, which is consistent with Bower (1962). However, there was also a tendency for the shifted groups to continue to perform faster on the postshift trials than the control groups that had consistently received the large GB_1 magnitude. Although there are no comparable results in the double runway (recall that Bower did not have control groups), similar findings have been obtained in studies concerned with the effects of drive shift on performance (Davis, 1957; Deese & Carpenter, 1951) where groups shifted from high to low drive prior to the performance of an instrumental task continued to perform faster than groups trained and tested at low-drive levels.

For the frustration theory, it is simply the case that an *increase* in the amount of reward in GB_1 would not result in frustration; therefore, the shifted groups would have been predicted to perform at the same rate in A_2 as the nonshifted control groups on postshift trials. For the perseverative theories, a small reward in GB_2 would be assumed to result in little instigation of r_G or T for Group S:S over the series of preshift trials as inferred from the preshift A_1 performance level. The shift to a larger reward in GB_1 in Group S-L:S would then be expected to reduce the strength of r_G or T below that of the control Group L-L:S in A_2. This should have resulted in *slower* performance in A_2 by Group S-L:S than Group L-L:S on postshift trials. In the case of Group S:L, preshift performance data indicate that instigation in A_1 due to the large reward in GB_2 brought r_G or T to the same level as Group L:L. A shift to a large reward in GB_1 for Group S-L:L would then be assumed to have brought r_G or T for this group to the same level as Group L-L:L in A_2. This should have resulted in equal postshift A_2 performance rates for these groups.

Neither theoretical approach could fully account for the data in the present study without additional assumptions concerning the effect of successive rewards under constant conditions, especially the effects that result from conditions of constant but unequal reward magnitudes in GB_1 and GB_2. In the case of frustration theory, the problem arises from the lack of unequivocal evidence that unequal reward magnitudes in GB_1 and GB_2 produce frustration to the small reward. The difficulty with perseverative theory is in determining more precisely the degree to which response strength is instigated in A_1 and the amount that perseverates into A_2. It would seem that further empirical as well as theoretical clarity will require additional studies of reward increase as well as reward decrease and the systematic variation of second as well as first goal compartment reward magnitude.

REFERENCES

Amsel, A., & Roussel, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *Journal of Experimental Psychology,* 1952, **43**, 363–368.

Atkinson, J. W. *An introduction to motivation.* New York: Van Nostrand, 1964.

Barrett, P., Peyser, C., & McHose, J. Effects of complete and incomplete reward reduction on a subsequent response. *Psychonomic Science,* 1965, **3**, 277–278.

Birch, J. D. Shift in activity and the concept of persisting tendency. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968.

Bower, G. The influence of graded reductions in reward and prior frustrating events upon the magnitude of the frustration effect. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 582–587.

Daly, H. Excitatory and inhibitory effects of complete and incomplete reward reduction in the double runway. *Journal of Experimental Psychology,* 1968, **76**, 430–438.

Davis, R. The effect of drive reversal on latency, amplitude, and activity level. *Journal of Experimental Psychology,* 1957, **53**, 310–315.

Deese, J., & Carpenter, J. Drive level and reinforcement. *Journal of Experimental Psychology,* 1951, **42**, 236–238.

Hays, W. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

McHose, J. Effect of continued non-reinforcement on the frustration effect. *Journal of Experimental Psychology,* 1963, **65**, 444–450.

Peckham, R., Jr., & Amsel, A. Magnitude of reward and the frustration effect in a within-subjects design. *Psychonomic Science,* 1964, **1**, 285–286.

Pubols, B. H., Jr. Incentive magnitude, learning, and performance in animals. *Psychological Bulletin,* 1960, **57**, 89–115.

Wagner, A. R. The role of reinforcement and non-reinforcement in an "apparent frustration effect." *Journal of Experimental Psychology,* 1959, **57**, 130–136.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received January 27, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 88–95

EFFECTS OF THE INTENSITY OF AUDITORY AND VISUAL READY SIGNALS ON SIMPLE REACTION TIME

DAVID L. KOHFELD[1]

United States Army Medical Research Laboratory, Fort Knox, Kentucky

The effects of auditory (Exp. I) and visual (Exp. II) ready signal intensity were investigated in a simple reaction time (RT) task. Mean RT to three auditory response signals was found to systematically increase with a corresponding increase in the intensity of either auditory or visual ready signals. The results were analyzed according to a decision model of stimulus intensity effects. It was concluded that ready signal intensity influenced the value of the detection criterion. Practice effects and individual differences were also significant determinants of the criterion level.

Previous investigations of simple reaction time (RT) have demonstrated that the stimulus context in which an RT signal is presented is an important factor in determining the speed of response. For example, Grice and Hunter (1964) reported larger intensity effects when two signals were presented randomly to the same group than when each signal was presented to a separate group. That is, contextual effects were apparent when *S*s received more than one response signal intensity per session and comparisons were within groups. The role of context was studied further by Murray and Kohfeld (1965) and Kohfeld (1968), who found that preadapting *S*s to various stimuli just prior to participation in an RT experiment significantly modified the resulting intensity functions in accordance with the average intensity of the preadaptation stimulus. Experimental manipulation of other variables, such as the temporal interval between signals (Bevan, 1968), the distribution of intensity values (Murray, 1969), and the intensity of the ready signal (Kohfeld, 1969), also indicates that RT is significantly affected by the relationship of a response signal to previous experimental events.

One manner of conceptualizing the contextual phenomena observed in RT situations is in terms of adaptation level (AL) theory. Since the theory predicts that the excitatory strength of a stimulus depends on its distance from AL, Grice and Hunter (1964) accounted for large within-*S* intensity effects by assuming that exposure to two intensities of test stimuli produced an AL between the two values, whereas exposure to only one intensity resulted in an AL at that particular value. Other studies (Kohfeld, 1968, 1969; Murray & Kohfeld, 1965) were also consistent with AL theory, as these experiments collectively indicated that both preadaptation stimuli and RT ready signals contributed to the effective stimulus context in which the response signals were presented.

In spite of the successes of AL theory in accounting for a variety of contextual phenomena (see Bevan, 1968), some writers contend that AL theory does not contain a specific principle of response evocation (Grice, 1968; Murray, 1969). These writers suggest that certain contextual variables (e.g., within-*S* presentation of stimuli and preadaptation procedures) can appropriately be viewed as influencing the detection criterion, a concept derived from McGill's (1961, 1963) decision-theoretical approach to stochastic latency mechanisms. McGill has postulated that the onset of an RT signal initiates a sequence of neural events that are accumulated over time. The *S* will respond to the signal when the cumulative impulse count reaches some predetermined number corresponding to the value of his detection criterion. The latency of *S*'s response is a measure of the time required for the impulse count to reach the criterion value. Since the impulse rate is determined by stimulus intensity and the impulse number is determined by the value of the detection criterion, the model predicts that RT will vary as a function of stimulus intensity and criterion level, respectively. Thus, any experimental manipulation that raises or lowers the criterion will have a corresponding effect on RT. It is in this latter respect that the concepts of AL and detection criterion are analogous as both theoretical approaches often make similar predictions regarding the effects of contextual variables on reaction latency.

---

[1] Requests for reprints should be sent to David L. Kohfeld, who is now at Southern Illinois University, Department of Psychology, Edwardsville, Illinois 62025.

The two experiments reported here dealt with the effects of ready signal intensity on RT. The primary aims of the research were twofold. First, an attempt was made to apply the decision-theoretical approach to the results. Based on the findings of previous studies (Baumeister, Dugas, & Erdberg, 1967; Kohfeld, 1969), it was hypothesized that high-intensity ready signals would produce longer mean RT than lower-intensity ready signals. If McGill's (1961, 1963) model is applicable, the hypothesized differences in RT offer evidence for the assumption that variations in ready signal intensity produce corresponding changes in the value of the detection criterion. The second purpose of the present research was to investigate whether cross-modal presentation of ready and response signals (visual–auditory) would produce effects similar to those obtained when both signals were in the same modality (auditory–auditory). Since the visual ready signal magnitudes employed in Exp. II were equated with the auditory ready signals presented in Exp. I, it was possible to determine whether modality was the limiting factor in the identification of ready signal effects.

EXPERIMENT I

Method

Subjects.—The *S*s were 16 soldiers, who were assigned to the laboratory after completion of basic training. A medical examination verified their physical health, and all *S*s were free of visual and hearing defects. The *S*s were assigned randomly to the various experimental conditions.

Apparatus.—The experiment was conducted in a double-wall, sound-treated chamber, which was totally dark during the entire RT session. The *S*s were seated in a desk chair, which had a conventional telegraph key clamped on its arm. The ready and response signals were three 1,000-cycle tones having intensities of 30-, 60-, and 90-db. SPL. The tones were generated by an audio oscillator and after appropriate attenuation were presented through calibrated earphones by means of an electronic switch with a rise and decay time of 10 msec. The durations of the ready and response signals were .5 and 1.5 sec., respectively. Foreperiods of 1, 2, or 3 sec. were given in irregular order on successive trials. There was a 15-sec. interval between response signal offset and ready signal onset. Timing of events was regulated by interval timers operating in a repetitive sequence. A paper tape reader and a system of shielded relays were used to select the foreperiod and tonal stimuli according to a programmed sequence. RT was registered in milliseconds by an electronic counter.

Procedure.—Each *S* participated in five consecutive RT sessions, one 45-min. session per day. Prior to Session 1, *S* was read conventional RT instructions. Each session was preceded by a 10-min. period of dark adaptation during which *S* sat quietly in the darkened sound booth. The *S* was then instructed over an intercom to adjust his earphones and place his right index finger on the telegraph key. Fifteen unscored practice trials ensued during which *S* became familiar with pressing the key as fast as possible to the second of two successive tones. A total of 90 scored trials per session was presented. A short rest period was given between each block of 30 trials. During each session, *S*s were administered the same order of 30-, 60-, and 90-db. response signals presented in random order, with the restriction that there were 10 presentations of each signal in each block.

Four ready signal conditions were presented to the 16 *S*s in a counterbalanced order on the first 4 consecutive days. For three of the conditions, the same ready signal (30 db., 60 db., or 90 db.) was presented throughout the entire session. The fourth condition (random) involved presenting each ready signal an equal number of times within each block of 30 trials. On Day 5, all *S*s were given the same condition (feedback) in which a 30-db. ready signal was given on all 90 trials and feedback of RT was reported over the intercom after each trial.

Results and Discussion

The data were analyzed according to an Orders × Conditions × Intensities analysis of variance. Significant sources of variation were attributed to response signal intensities, $F\ (2, 24) = 130.79$, $p < .001$, ready signal conditions, $F\ (4, 48) = 17.52$, $p < .001$, and to the Conditions × Intensities interaction, $F\ (8, 96) = 2.55$, $p < .025$. The large effect due to response signal intensity replicated the common finding that RT undergoes a systematic decrease with a corresponding increase in stimulus intensity. The absence of an effect due to orders ($F = .19$) indicated that counterbalancing procedures were statistically adequate. Mean RTs in milliseconds for the five ready signal conditions were as follows: 90 db., 304; random, 289; 60 db., 284; 30 db., 267; and feedback, 234. In view of the significant conditions effect, the hypothesis that RT is related to ready signal intensity was confirmed. It is of additional interest to note the similarity in mean RTs for the 60-db. and random conditions. Further inspection of the data revealed that *S*s in the random condition did not respond differentially to the 30-, 60-, and 90-db. ready signal trials. This finding is consistent with the results of a previous study (Kohfeld, 1969) in which random presentation of ready signals produced RTs that were similar to those obtained when a single ready signal value at the mean of the stimuli was presented. When ready signals are presented in an unpredictable order, it appears that *S*s maintain an effective reference level which is intermediate among the intensity values. Finally, the significant Conditions × Intensities interaction is meaningful when viewed in conjunction with the finding that RT was shortest for the feedback condition. Assuming that the ready signal conditions influenced the detection criterion, these results imply that intensity effects were smaller for a low criterion value (feedback condition) than for a relatively high criterion value (90-db. condition). This implication is discussed subsequently more fully.

Figure 1 presents the results in accordance with a method described by Grice (1968) in his evaluation of the decision theory model. As noted earlier, the model assumes that the number of impulses required for a response is determined by the value of the detection criterion. The hypothetical impulse dimension in Fig. 1 serves as an estimate of the relative positions of the criteria. The rationale for determining these positions has been discussed elsewhere (Grice, 1968, pp. 368–371). Briefly, four steps were followed: (*a*) The grand means for each of the five ready signal conditions were calculated; (*b*) 100 msec. were subtracted from each of the means in order to account for the assumed value of the "irreducible minimum" (Woodworth & Schlosberg, 1954); (*c*) the cri-

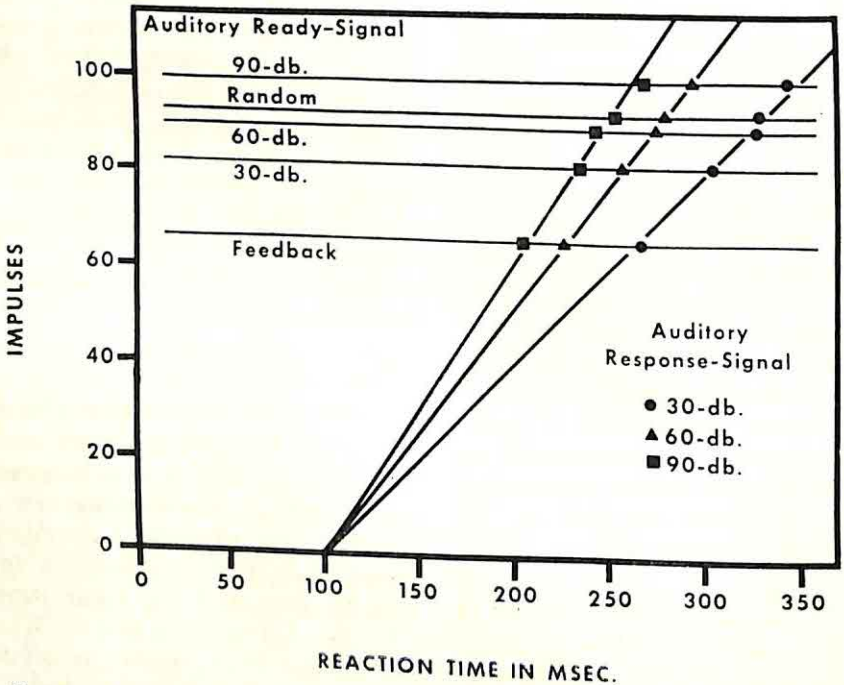


FIG. 1. Decision model applied to the effects of auditory ready signal conditions and response signal intensity on reaction time.

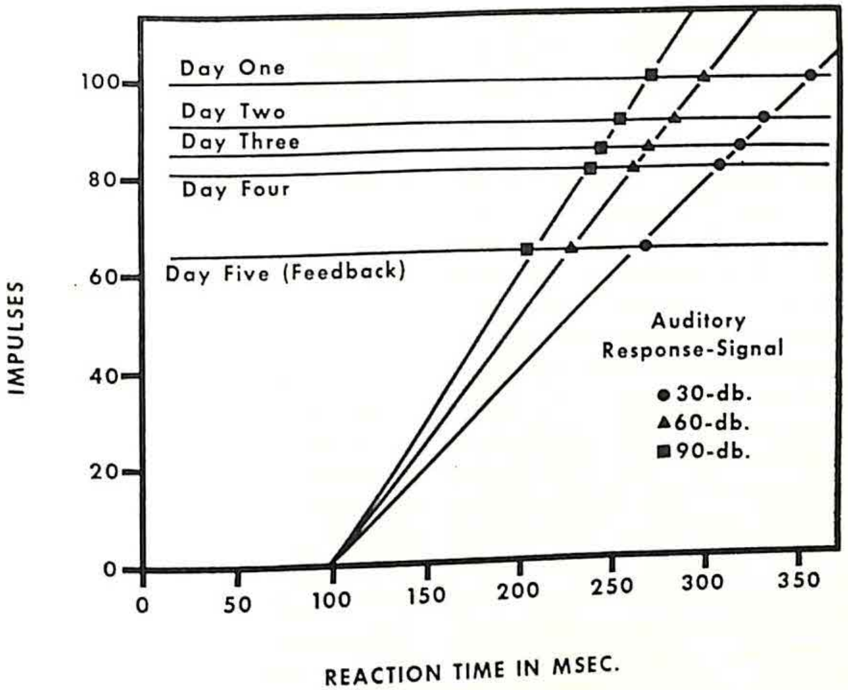


FIG. 2. Decision model applied to the effects of practice (Days 1–5) and response signal intensity on reaction time.

terion for the 90-db. condition was arbitrarily assigned the index value of 100 impulses; and (*d*) using the values obtained in *b*, the other ready signal criteria were calculated as a proportion of the 90-db. index value. The 15 points in Fig. 1 correspond to the mean RTs for each of the five RT sessions. The points were plotted along the horizontal criterion lines representing the five ready signal conditions. The three linear functions represent the rates at which the impulses were assumed to accumulate for the response signal inputs. The slopes of the inputs were estimated by least squares fits to the five means for each of the response signal intensities. Slope values for the inputs were subject to two restrictions: (*a*) The position of the criterion lines was based on the arbitrarily assigned index value of 100 impulses for the 90-db. condition; and (*b*) each function originated at 100 msec. on the time axis because of the assumed irreducible minimum, which consists of RT components not under the influence of stimulus intensity. Thus, the observed mean RTs are the points on the time axis corresponding to the intersections of the input functions with the appropriate criterion value. While tests for goodness of fit were not made, it is apparent from Fig. 1 that the data points deviate only slightly from the estimated intersections of the input functions with the criteria. Since the model assumes that the input functions fan out from the common origin of 100 msec., larger intensity effects were expected for relatively high criterion values than for lower values. As noted earlier, this prediction received support from the significant Conditions × Intensities interaction. It was concluded from the present analysis that the decision theory model provides one way to account for the effects of ready signal intensity.

Grice (1968) has suggested that the practice effect typically observed in RT experiments may result from a progressive lowering of the detection criterion. He found that RT data obtained on 2 consecutive days, when presented in terms of the decision model, supported the assumption that *S*s' criterion decreased on Day 2. The data from the present experiment were particularly suitable for further evaluation of this hypothesis, as RT measures from 5 consecutive days of testing were available. Since the ready signal conditions were counterbalanced over the first 4 days, the data from Day 5 were excluded from statistical analyses. A Days × Conditions × Intensities × *S*s analysis of variance indicated the following significant effects: days, $F\ (3, 9) = 17.24,\ p < .001$; conditions, $F\ (3, 9) = 11.01,\ p < .001$; intensities, $F\ (2, 6) = 50.61,\ p < .001$; and Days × Intensities, $F\ (6, 18) = 6.03,\ p < .005$. In view of the nonsignificant Days × Conditions interaction ($F = 1.87$), the ready signal means within each day were pooled into daily grand means. For purposes of graphic presentation, the data from Day 5 also were included among the following daily means: Day 1, 308; Day 2, 289; Day 3, 276; Day 4, 269; and Day 5, 234. In a manner similar to that described previously, the five daily means were used to obtain estimates of the criterion values for Days 1–5. The input functions were also fitted in the same manner as before. This analysis is presented in Fig. 2. It is obvious that the model accounts for nearly all of the variance among the data points. Furthermore, larger intensity effects were obtained for relatively high criterion values than for lower values of the criterion, a finding that received support from the significant Days × Intensities interaction. The data are consistent with the interpretation that one effect of practice in RT tasks is a progressive lowering of the detection criterion.

An important feature of most RT experiments is the wide and consistent individual differences among the mean RTs of various individuals. In terms of the decision model, it seems possible that individual differences in performance may reflect characteristic differences in the detection criteria of various individuals. In other words, *S*s with longer mean RTs may adopt a relatively higher or more conservative criterion than *S*s with shorter mean RTs. In order to evaluate this hypothesis, the 16 *S*s in the present sample were divided into eight fast and eight slow responders as indicated by their grand mean RT scores. The data were then analyzed according to a Fast vs. Slow × Conditions × Intensities analysis of variance. Significant sources of variation were attributed to fast vs. slow *S*s, $F\ (1, 14) = 13.47,\ p < .005$; ready signal conditions, $F\ (4, 56) = 17.50,\ p < .001$; response signal intensities, $F\ (2, 28) = 245.33,\ p < .001$; and to the Fast vs. Slow × Intensities interaction, $F\ (2, 28) = 12.41,\ p < .001$. Since the Fast vs. Slow × Conditions interaction was not significant ($F = .62$), the ready sig-

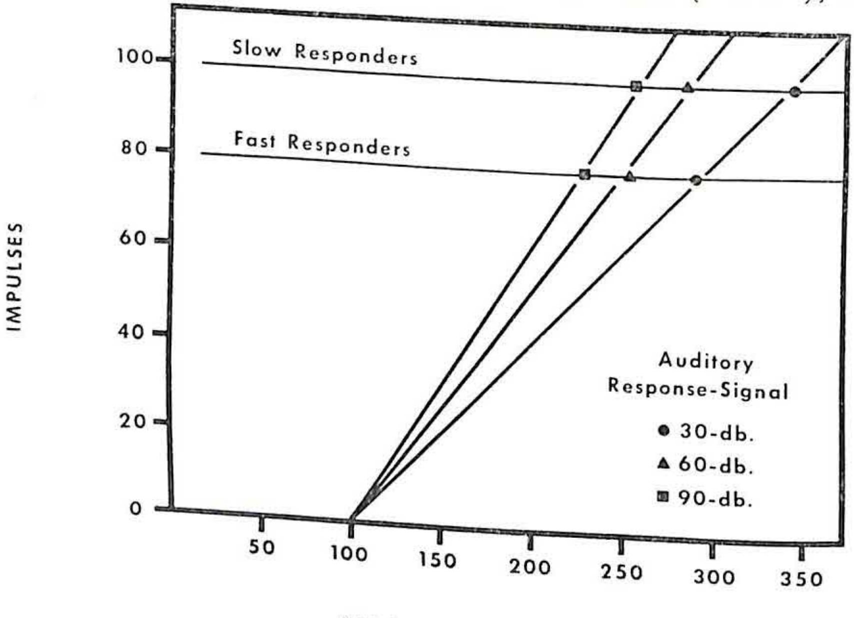


FIG. 3. Decision model applied to the effects of individual differences (fast vs. slow responders) and response signal intensity on reaction time.

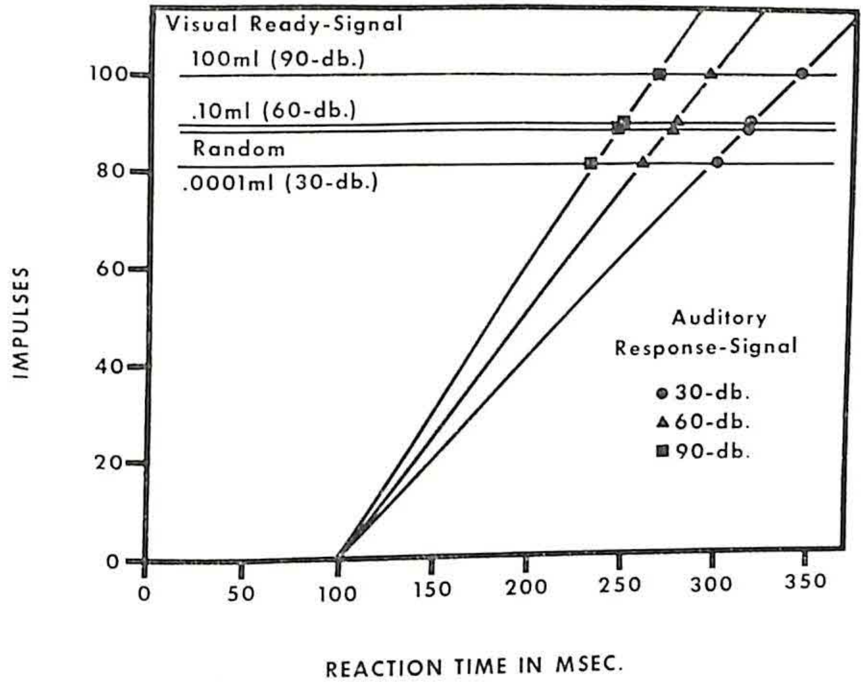


FIG. 4. Decision model applied to the effects of visual ready signal conditions and auditory response signal intensity on reaction time.

nal means were pooled into grand means for the slow and fast responders. Figure 3 presents the data in accordance with the method outlined previously. Note again how the three linear input functions originate at 100 msec. and pass through the grand means for the two groups. The significant Fast vs. Slow × Intensities interaction supports the model's prediction that the slope of the RT intensity function is steeper for a relatively high criterion value. In terms of the decision model, the results support the hypothesis that individual variation in mean RT reflects corresponding variation in the values of Ss' detection criteria.

EXPERIMENT II

Method

Subjects and procedure.—The Ss were 16 soldiers who were selected in the same manner as in Exp. I. The procedure was essentially the same as in Exp. I, with two exceptions: (*a*) A visual ready signal was used; and (*b*) Day 5 (feedback condition) was omitted because of scheduling difficulties.

Apparatus.—The apparatus was the same as that used in Exp. I except that visual ready signals were programmed in the trial-to-trial sequence. The ready signals were three intensities of white light presented to the dark-adapted *S* through a 15 × 15 in. window of ⅛-in. milky Plexiglas. The window was situated in the front panel of a light box, which was mounted at eye level approximately 3 ft. from *S*. During RT trials, *S* was instructed to look straight ahead and keep his eyes open. The three light levels were set at 100, .10, and .0001 mL. as calibrated by a dark-adapted *O* with a Macbeth illuminometer. These light levels were chosen because they correspond in subjective magnitude to 90-db., 60-db., and 30-db. tones, respectively (Stevens, Mack, & Stevens, 1960). In accordance with Stevens' (1955) suggestion that a decibel scale is appropriate for both sound intensity and light intensity, the ready signals employed in Exp. II are referred to as 30-db. (.0001 mL.), 60-db. (.10 mL.), and 90-db. (100 mL.) light levels.

Results and Discussion

Figure 4 presents the results in accordance with the decision model. The procedures for positioning the detection criteria and fitting the input functions were identical to those described in Exp. I. Of particular interest was the similarity between the results of Exp. II and those of Exp. I. That is, mean RT was shortest for the 30-db. ready signal, longest for the 90-db. ready signal, and intermediate for the 60-db. and random conditions. An analysis of variance indicated that the effects of ready signal intensity were significant, $F\ (3, 36) = 4.14$, $p < .025$. In addition, the effects of practice and individual differences were analyzed in the same manner as in Exp. I. While not presented here, the results of these analyses were consistent with those presented in Fig. 2 and 3. In general, the results of Exp. II indicated that ready signal effects were not restricted to a single modality; rather, psychophysically matched visual and auditory ready signals had similar effects on the RT intensity functions.

One difference in the findings of the two experiments is worth mentioning. In Exp. II, the Ready Signal Conditions × Response Signal Intensities interaction was not statistically significant ($F = .41$). This appears to contradict the model's assumption that larger intensity effects result from relatively high criterion values than from lower values of the criterion. Moreover, two other experiments also resulted in nonsignificant Stimulus Intensity × Adaptation (Criterion) Conditions interactions (Kohfeld, 1968; Murray & Kohfeld, 1965). While the data from Exp. II appear to be consistent with the model, as shown in Fig. 4, the data from Exp. I suggest that relatively wide manipulation of the criterion (Day 5, feedback condition) is necessary in order to statistically confirm the predicted interaction.

Discussion

The main finding of the present experiments was that mean RT systematically increased with a corresponding increase in ready signal intensity. A major purpose of the research was to relate this finding to McGill's (1963) and Grice's (1968) decision-theoretical approach to response evocation. As described previously, the decision model assumes that the rate at which neural impulses accumulate is determined by the intensity of a stimulus input, whereas the number of impulses required for a response is determined by the value of the detection criterion. The present results are consistent with the model if one assumes that response signal intensity determined the slope of the input function, whereas ready signal intensity influenced the value of the detection criterion.

It is noteworthy that McGill's (McGill & Gibbon, 1965) primary concern has been to deduce mathematically a multistage process that adequately describes various input functions. Since he assumes that the criterion remains relatively stable over trials, trial-to-trial variation in RT is attributed to the variable rate at which neural impulses accumulate. On the other hand, Grice (1968) has argued that variability in reaction latency can be attributed more appropriately to fluctuations in *S*'s detection criterion. Since the criterion appears to be influenced readily by a variety of experimental and individual difference variables, Grice's position has theoretical value. The present findings, e.g., indicated that ready signal intensity, degree of practice, and individual differences were significant determinants of RT and, presumably, *S*'s detection criterion. While the present research leaned heavily on McGill's assumptions regarding sensory inputs, it is clear that the primary emphasis was on the role of criterion variation in reaction latency.

The present results appear to be consistent with an adaptation level approach to contextual phenomena. AL theory states that the experience of a given intensity can be described in terms of its contrast with some reference level of stimulation. If one assumes that the ready signal served as a reference against which the response signals were compared, the 90-db. ready signal produced the longest mean RT because the response signals were all at or below AL, whereas the 30-db. ready signals resulted in shortest mean RT, the response signals all being at or above AL. Similarly, the 60-db. and random conditions produced intermediate levels of responding, the effective AL being close to the mean of the response signals. While it appears that the concepts of AL and detection criterion have analogous properties, the latter concept provides a more specific description of the mechanism of response evocation, at least for studies of stimulus intensity effects in RT (Grice, 1968). Furthermore, Murray (1969) has pointed out that AL theory provides no way to account for the fact that relatively high criterion values (high AL) produce larger stimulus intensity effects than when lower criterion values (lower ALs) are assumed. The decision model, on the other hand, predicts a Signal Intensity × Criterion Level interaction, a prediction supported by Murray's results and by those reported in Exp. I.

Another advantage of the decision model is that it may account for Behar and Adams' (1966) finding that RT was inversely related to ready signal intensity, a result that contrasts with the present findings and with those of previous RT studies (Baumeister et al., 1967; Kohfeld, 1969). Three aspects of Behar and

Adams' study are worth mentioning. First, they employed a conditioning-like methodology in which only one response signal was presented on all trials, and little within-session variation in foreperiod interval was used. Second, the ready signal overlapped temporally with the response signal, both signals being terminated simultaneously. Third, and perhaps most noteworthy, they concluded that the ready signal in their experiments had properties analogous to those of the CS in classical conditioning. In this vein, it is noteworthy that Grice, Hunter, Kohfeld, and Masters (1967) reported that a relatively intense CS produced shorter latency CRs than a weaker CS. The analogous nature of Behar and Adams' and Grice et al.'s results is especially significant in view of recent proposals that classical conditioning variables can be studied within the context of decision theory (Grice, 1968; Suboski, 1967). For example, Grice (1968, p. 362) has suggested that the intensity of the CS in classical eyelid conditioning determines the slope of the input function. With respect to Behar and Adams' RT design, if one assumes that ready signal intensity determined the rate of impulse accumulation, whereas other variables (e.g., trace vs. delayed presentation) influenced *S*'s detection criterion, then the results are not only consistent with the decision model, but also support their conclusion that the ready signal behaved like a CS.

In Exp. II, it was found that visual ready signals influenced the detection criterion in the same manner as when auditory ready signals were used. This finding has at least two implications. First, it demonstrates that ready signal effects, and consequently the criterion, are not restricted to peripheral mechanisms, but reflect a process that could be termed central in nature. The second implication of Exp. II is that the concept of a detection criterion may provide an explanation for intersensory effects, at least for cases in which an interval of 1 sec. or longer separates the stimuli. Intersensory effects are said to occur when responses are either facilitated or inhibited by the presentation of stimuli in two or more modalities. The fact that the latency of a response to an auditory stimulus depended on the intensity of a preceding visual stimulus, along with the inference that this effect was mediated by criterion variation, suggests that future study of sensory interaction may profit from a decision-theoretical approach.

REFERENCES

Baumeister, A. A., Dugas, J., & Erdberg, P. Effects of warning signal intensity, reaction signal intensity, preparatory interval, and temporal uncertainty on reaction times of mental defectives. *Psychological Record,* 1967, **17**, 503–507.

Behar, I., & Adams, C. K. Some properties of the reaction-time ready-signal. *American Journal of Psychology,* 1966, **79**, 419–426.

Bevan, W. The contextual basis of behavior. *American Psychologist,* 1968, **23**, 701–714.

Grice, G. R. Stimulus intensity and response evocation. *Psychological Review,* 1968, **75**, 359–373.

Grice, G. R., & Hunter, J. J. Stimulus intensity effects depend upon the type of experimental design. *Psychological Review,* 1964, **71**, 247–256.

Grice, G. R., Hunter, J. J., Kohfeld, D. L., & Masters, L. Order of presentation, CS intensity, and response latency. *Journal of Experimental Psychology,* 1967, **74**, 581–585.

Kohfeld, D. L. Stimulus intensity and adaptation level as determinants of simple reaction time. *Journal of Experimental Psychology,* 1968, **76**, 468–473.

Kohfeld, D. L. Effects of ready-signal intensity and intensity of the preceding response-signal on simple reaction time. *American Journal of Psychology,* 1969, **82**, 104–110.

McGill, W. J. Loudness and reaction time. *Acta Psychologica,* 1961, **19**, 193–199.

McGill, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* Vol. 1. New York: Wiley, 1963.

McGill, W. J., & Gibbon, J. The general-gamma distribution and reaction times. *Journal of Mathematical Psychology,* 1965, **2**, 1–18.

Murray, H. G. Effects of stimulus intensity and method of stimulus presentation on simple reaction time: Evaluation of a decision-theory model. (Doctoral dissertation, University of Illinois) Ann Arbor, Mich.: University Microfilms, 1969. No. 69-1397.

Murray, H. G., & Kohfeld, D. L. Role of adaptation level in stimulus intensity dynamism. *Psychonomic Science,* 1965, **3**, 439–440.

Stevens, J. C., Mack, J. D., & Stevens, S. S. Growth of sensation on seven continua as measured by force of handgrip. *Journal of Experimental Psychology,* 1960, **59**, 60–67.

Stevens, S. S. Decibels of light and sound. *Physics Today,* 1955, **8**, 12–17.

Suboski, M. D. The analysis of classical discrimination conditioning experiments. *Psychological Bulletin,* 1967, **68**, 235–242.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, Rinehart & Winston, 1954.

(Received January 31, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 96–106

FORGETTING IN SHORT-TERM RECALL: ALL-OR-NONE OR DECREMENTAL?

THOMAS O. NELSON[1] AND WILLIAM H. BATCHELDER
University of Illinois

Three experiments employing various replacement-repetition methods investigated the nature of the forgetting process in the Peterson and Peterson short-term memory (STM) paradigm. In the first two experiments, incorrectly recalled CCCs were repeated and their subsequent recall was compared with the recall of new CCCs. In the third experiment, several schemes for replacing individual letters prior to repetition were compared. Recall following relearning was superior to recall following new learning only when some of the letters of the repeated CCCs had been correct prior to repetition. Significant numbers of phonemic confusions occurred, but had little value for predicting relearning. It was concluded that (*a*) letters are forgotten in an all-or-none fashion and (*b*) benefits of relearning over original learning occur only when segments of the relearned material do not, in fact, have to be relearned.

Following Rock's (1957) study, a flurry of experiments probed into the all-or-none vs. incremental learning controversy. Tulving (1964) concluded that learning is neither all-or-none nor incremental, but rather, simply, *all*. If this proposition is valid, then incorrect recall is the result of forgetting, not incomplete original learning (cf. Bregman, 1967). The present study investigated whether forgetting, in particular that following short-term memory (STM), is an all-or-none or decremental process.

Decremental forgetting stands in natural symmetry to the Hull-Spence theory of incremental learning. Gradual reduction in the memory trace results in forgetting when the trace goes below the recall threshold, and additional loss of subthreshold associative strength is gradual. Put differently, the decremental process suggests a continuous multistate forgetting model (see Fig. 1).

One all-or-none explanation of forgetting suggests that an item, upon being forgotten, immediately loses all of its associative strength and recedes back to the unlearned (U) state (Bernbach, 1965; Bower, 1962; Calfee & Atkinson, 1965). Another all-or-none view, suggested by Atkinson and Crothers (1964, p. 288), is that when an item is forgotten, it does not return to the U state; instead, it will recede only as far as a forgotten (F) state, which in associative strength terminology is somewhere between the recall threshold and the U state. Although *S* guesses in both states, relearning is more likely from State F. Figure 1 compares these three conceptions of the forgetting process.

More recently, several investigators have represented STM by a finite-slot storage buffer (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). The implication of such models is that an item either is or is not in STM; however, these formulations do not imply necessarily that individual items can be represented by a simple two-state all-or-none model since the probability that an item is stored on presentation may depend on the current state of the entire storage buffer.

Data related to these ideas can be obtained by using the Peterson and Peterson (1959) STM paradigm. In this paradigm, it is assumed that every item is learned during acquisition because immediate retention tests yield nearly 100% correct recall; thus, "one-trial acquisition of these simple items is assured, although retention is a function of the interval between presentation and recall [Bernbach, 1965, p. 128]." If a forgotten item retains some subthreshold associative strength (as in the decremental model and F-and-U model), it should be easier to recall after being re-presented than items presented

---

[1] Requests for reprints should be sent to Thomas O. Nelson, Department of Psychology, University of Illinois, Urbana, Illinois 61801.

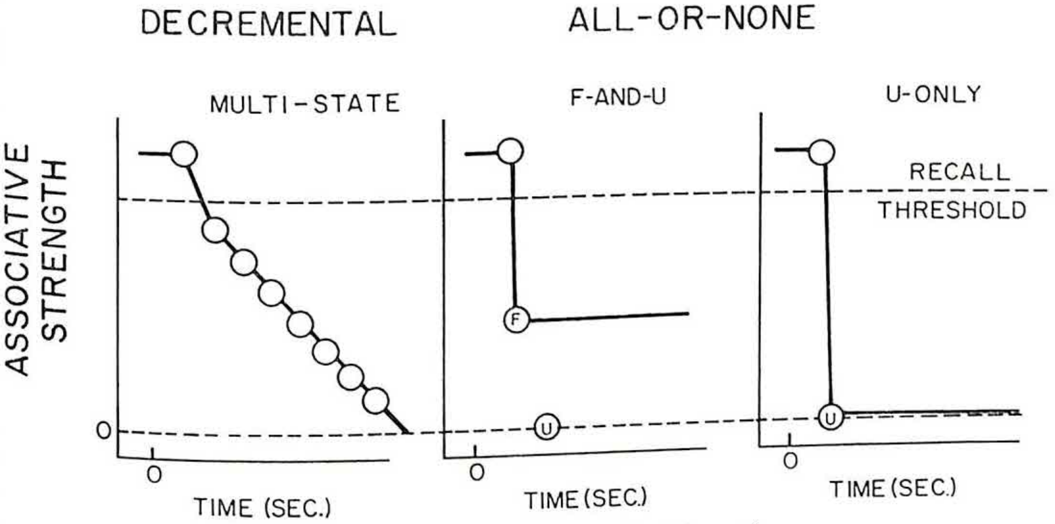


FIG. 1. Comparison of three models of forgetting.

for the first time; if a forgotten item effectively loses all of its associative strength (as in the U-only model), it should not be any easier to recall after relearning than items learned for the first time.

EXPERIMENT I

In Exp. I, *S*s were given an initial block of six CCC items. Incorrect items were repeated and correct items were replaced in the second block of six trials. The intention was to compare recall of re-presented forgotten items with that of new items.

Method

Subjects and materials.—The *S*s were 34 undergraduate students from the University of Illinois, each run individually. Eleven CCC items that had association values of 21–25% were selected (Witmer, 1935). The items were presented with a Kodak Carousel projector; a California Industries Model 220A tape reader triggered the advancement of each slide.

Procedure and instructions.—Each *S* received a sequence of 14 Peterson and Peterson trials (which included 2 practice trials). The presentation order for the four slides comprising a trial was: "Ready" signal (2 sec.), CCC item (1 sec.), retention interval (10 sec.), and recall interval (10 sec.). The *S* read each CCC item aloud during acquisition. During the retention interval a slide with five rows of 12 random digits was presented; *S* read the digits aloud as fast as possible. The signal for recall was an asterisk; *S* was instructed to recall always three letters in original order, even if he had to guess.

After *S* was instructed in the procedure, two practice trials with nonexperimental items were run. If *S* had no questions, the first block of six trials began. If *S* was incorrect on either all or none of the items, he was eliminated from the experiment (resulting in the loss of two *S*s). An item was considered correct only if *S* recalled all three letters in their original serial positions. For *S*s who forgot at least one, but no more than five, items, the second block of six trials began without interruption.

The items in the second block were chosen according to two criteria: (*a*) Items that were incorrect during the first block were re-presented in the same sequential position in the second block; thus, there were always 5 items intervening between the first and second presentations of an incorrectly recalled item. (*b*) Items that were correct during the first block were replaced by new items in the second block. The 11 items selected for the experiment were used approximately equally often as Block 1 items and Block 2 replacement items.

Results

Item-level analyses.—First, the data were scored as errors or successes in recall of the entire CCC item. Each *S* donated three dependent measures: the proportion correct during the first block (B_1), the proportion correct on new (replacement) items during the second block (N_2), and the proportion correct on repeated items during the second block (R_2). Each of *S*'s three proportion scores was converted into radians ($\phi = 2$ arc sine P; Snedecor, 1956) because the denominators varied across *S*s in accord with the number of Block 1 errors. It should be noted that analyses involving the untransformed proportions were in essential agreement with those of the transformed proportions.

To compare Block 2 performance on repeated and replacement items, a *t* test for paired scores was performed on R_2 and N_2. The results yielded only marginal evidence for improved recall for repeated items, $t(31) = 1.90$, $p < .10$. Comparing B_1 with N_2 failed to reveal any evidence for a change in performance on new items between Blocks 1 and 2, $t(31) = .56$.

In response-contingent replacement studies, individual-difference factors can confound the results. Table 1 presents the intercorrelation matrix for B_1, N_2, and R_2. If individual differences were operating, high correlations would be expected. Each correlation is based on 32 cases, and none of them approaches statistical significance.

Analyses based on number of letters missed originally.—Viewed in the light of the research on the relationship between the number of presented chunks and STM (Melton, 1963; Murdock, 1961), it seems naive to pose the question of the nature of the memory trace solely on the level of the entire CCC item. Therefore, the proportion of correct repeated items was assessed as a function of the number of letters missed during the original B_1 trial. First, the data were pooled over *S*s and items. Then, the conditional proportion of items correct on repetition given *i* letters incorrect during Block 1, $P(C|iW)$, was computed for $i = 1, 2, 3$. With $P(C|N)$ representing the proportion correct on new (Block 2) items and $P(C|R)$ the proportion correct on repeated items [i.e., $P(C|1W \text{ or } 2W \text{ or } 3W)$], the results of these analyses are presented in Table 2.

TABLE 1

INTERCORRELATION MATRIX FOR B_1, N_2, AND R_2 IN EXP. I

Measure	Measure		
	B_1	R_2	N_2
B_1	—	—	—
R_2	.089	—	—
N_2	.052	.062	—

TABLE 2

PROPORTION OF ITEMS CORRECT IN BLOCK 2, CONDITIONAL ON NUMBER OF LETTERS MISSED IN BLOCK 1

Measure	Proportion	Denominator
$P(C\|N)$	.405	84
$P(C\|3W)$	.429	21
$P(C\|2W)$	.549	51
$P(C\|1W)$	.611	36
$P(C\|R)$	.546	108

The pattern in Table 2 supports the notion that the more letters missed originally, the poorer the performance during re-presentation. Since the intercorrelation matrix in Table 1 indicates no evidence for individual differences, the pooled data were subjected to *z* tests of the difference between each error proportion on repeats and $P(C|N)$. Comparing $P(C|R)$ with $P(C|N)$ yields $z = 1.94$, $p = .05$, which is consistent with the results of the *t* test for repeats. The *z* scores for $P(C|iW) - P(C|N)$ are as follows: for $i = 1$, $z = 2.07$, $p = .04$; for $i = 2$, $z = 1.63$, $p = .10$; and for $i = 3$, $z = .20$, $p = .84$. These analyses suggest that the improved performance on repeated items is coming from items that *S* only partially failed to recall during Block 1.

Letter-level analyses.—Additional evidence for the role of individual letters in recall comes from an analysis of the recall of single letters. Let $P(c|n)$ be the proportion of letters correct for new items, $P(c|r)$ be the proportion of originally missed letters that were correct on repetition, and $P(c|iw)$ be the proportion of originally missed letters that were correct on repetition, given that *i* letters were missed originally. For example, $P(c|2w)$ is the proportion of letters correct

TABLE 3

PROPORTION OF ORIGINALLY INCORRECT LETTERS CORRECT ON REPETITION, CONDITIONAL ON NUMBER OF LETTERS ORIGINALLY MISSED

Measure	Proportion	Denominator
$P(c\|n)$	.571	252
$P(c\|3w)$	.619	63
$P(c\|2w)$	.608	102
$P(c\|1w)$	.778	36
$P(c\|r)$	.642	201

on repetition that were missed in items with two letters wrong during Block 1 trials. Table 3 presents these conditional proportions.

A z test comparing $P(\mathrm{c}|\mathrm{n})$ with each repetition proportion was performed. The results were as follows: with $P(\mathrm{c}|\mathrm{r})$, $z = 1.53$, $p = .13$; with $P(\mathrm{c}|1\mathrm{w})$, $z = 2.37$, $p = .02$; with $P(\mathrm{c}|2\mathrm{w})$, $z = 1.30$, $p = .19$; and with $P(\mathrm{c}|3\mathrm{w})$, $z = .34$, $p = .73$. The one significant difference indicates that repetition improved recall only when no more than one letter had been missed during Block 1. The near identity of $P(\mathrm{c}|\mathrm{n})$ and $P(\mathrm{c}|3\mathrm{w})$ is additional evidence that an item with three letters missed originally has no residual letter or item trace. The proportion of originally correct letters that remained correct in Block 2 was .790 ($N = 124$); a z test comparing this proportion with $P(\mathrm{c}|\mathrm{n})$ yielded $z = 4.13$, which indicates that previously correct letters have higher retention following re-presentation than do new letters presented for the first time.

Discussion

The results of Exp. I all point to an improved performance following re-presentation only if some letters were correct on original presentation. The improved performance on repeats could be due to remembered letters (or letter sequences) from Block 1, which may reduce the effective item size in Block 2. Even more facilitation from re-presentation might have occurred if the proportion of originally correct letters remaining correct after repetition had been higher than .790.

Melton (1963), taking theoretical account of Murdock's (1961) results, argues that the recall curve for an item depends on the number of letters (or chunks) in that item; thus, new (N_2) items could be following a steeper retention curve than the repeated items with facilitation from previous partial recall. If Melton's curves are simply due to long items having more separate letters to be remembered, the retention curves for a single letter in an item should be the same for various item lengths. If this is so, then the proportion of originally missed letters that become correct on repetition should not depend on original performance. However, the fact that Block 2 performance on a single previously missed letter is significantly higher than performance for new letters suggests that the effective length of the item to be learned (or relearned) may also influence single-letter retention. This would suggest that the short-term retention of a single chunk would depend on the number of chunks already in memory. This assumption is not inconsistent with Melton's demonstration that first-letter retention, as well as entire-item retention, is a function of item length.

EXPERIMENT II

The failure to find stronger evidence for the preceding propositions may have been due to the five items that intervened between a Block 1 item and its re-presentation. Even if there was facilitation from one or two previously recalled Block 1 letters, interference from the five interpolated items may have counteracted the facilitation from previously recalled letters. It may be that for a trace to strongly benefit relearning, the re-presented item must somehow "connect" with its own prior trace. In addition, the memory for missed letters may even have been a decremental process that reset to zero during the 127 sec. before Block 2 repetition began. Experiment II was performed to test the role of an interpolated interval (without intervening items) between original presentation and repetition as well as to provide a replication of Exp. I.

Method

The *S*s were 25 *S*s who took part in Exp. I, and they were run individually after a brief rest from Exp. I. The general procedure was the same as that for the previous experiment except for four modifications: (*a*) The 20 items were again four modifications: CCCs, but this time they had a Witmer association value between 29% and 33%, i.e., slightly easier items. (*b*) The *S*s were instructed to read the items silently to themselves during acquisition. (*c*) Each time an item was missed, it was re-presented either 5, 25, or 65 sec. later; this intertrial interval was filled with number reading. Correct items were also followed by the 5-, 25-, or 65-sec. intertrial interval, but they were not re-presented. (*d*) Each *S* received a total of 20 trials, including both the trials with new items and the trials with re-presented items; the order of item presentation varied randomly across *S*s. A tape reader programmed the intertrial intervals in a random order.

Results

Preliminary.—To test for homogeneity in item difficulty, a chi-square test was run on the number of times that each of the 20 items was correctly recalled during first presentation, χ^2 (19) = 14.7, $p > .50$. This indicated no evidence for item differences.

Intertrial interval.—The performance on repeated items was computed for each of the three intertrial intervals. The data were first pooled over *S*s and items, and then the proportion correct at each intertrial interval was assessed. These proportions were .67, .67, and .71 for intertrial intervals of 5, 25, and 65 sec., respectively. There is no evidence for an intertrial interval effect, although the denominators of 64, 30, and 68, respectively, preclude strong confidence in the identity of these proportions.

Item-level analyses.—Since the intertrial interval did not seem to be a factor, the data were pooled over intertrial intervals and analyzed like those from Exp. I. Each *S* in Exp. II donated two dependent measures: (*a*) the proportion correct on original items and (*b*) the proportion correct on repeated items. These data were transformed to radians, and a *t* test of paired scores yielded a t (24) = 5.18, $p < .001$, indicating strong evidence that performance on repeated items was superior to performance on original items.

The data were pooled also over *S*s and items and subjected to an analysis of the proportion of correct items conditional on the number of letters missed originally. Let $P(\mathrm{C}|\mathrm{O})$ be the proportion of correct original items, $P(\mathrm{C}|\mathrm{R})$ be the proportion of correct repeated items, and $P(\mathrm{C}|i\mathrm{W})$ be the proportion of correct repeated items conditional on i letters missed originally, $i = 1, 2, 3$. Table 4 presents these proportions, which provide a close replication of Table 2 (Exp. I) except that all of the present proportions are higher. Again, the proportion correct on repeats was a decreasing function of the number of letters missed initially. In addition, $P(\mathrm{C}|3\mathrm{W})$ was nearly identical to $P(\mathrm{C}|\mathrm{O})$, suggesting that if all three letters were missed originally, performance on representation is the same as original performance.

TABLE 4

Proportion of Items Correct, Conditional on Number of Letters Originally Missed

Measure	Proportion	Denominator
$P(\mathrm{C}\|\mathrm{O})$	.514	333
$P(\mathrm{C}\|3\mathrm{W})$	.519	27
$P(\mathrm{C}\|2\mathrm{W})$	.646	65
$P(\mathrm{C}\|1\mathrm{W})$	.786	70
$P(\mathrm{C}\|\mathrm{R})$	.685	162

TABLE 5

Proportion of Originally Incorrect Letters Correct on Repetition, Conditional on Number of Letters Originally Missed

Measure	Proportion	Denominator
$P(\mathrm{c}\|\mathrm{o})$	.716	999
$P(\mathrm{c}\|3\mathrm{w})$	.630	81
$P(\mathrm{c}\|2\mathrm{w})$	.773	130
$P(\mathrm{c}\|1\mathrm{w})$	.829	70
$P(\mathrm{c}\|\mathrm{r})$	.746	281

Separate z tests were done comparing $P(\mathrm{C}|\mathrm{O})$ with each of the repeat proportions. The results were: with $P(\mathrm{C}|\mathrm{R})$, $z = 3.61$, $p < .001$; with $i = 1$, $z = 4.16$, $p < .001$; with $i = 2$, $z = 1.95$, $p = .05$; and with $i = 3$, $z = .05$, $p = .96$. Thus, the evidence is quite strong that the significant improvement in repetition is due to items that had one, or at most two, letters missed originally.

Individual-letter analyses.—The recall process on individual letters provides further evidence that improved recall following representation is due to letters previously remembered. That previously remembered letters remained in memory during the intertrial interval is shown by the proportion of letters correct on Trial $n + 1$ given that they were previously correct on Trial n; this conditional proportion was .932.

Let $P(\mathrm{c}|\mathrm{o})$ be the proportion of letters correct originally, and let $P(\mathrm{c}|\mathrm{r})$ and $P(\mathrm{c}|i\mathrm{w})$ be defined as in the Exp. I individual-letter analysis. Table 5, which reports these proportions, supports the conclusions of the previous analysis. The difference between $P(\mathrm{c}|1\mathrm{w})$ and $P(\mathrm{c}|\mathrm{o})$ is the only significant comparison involving $P(\mathrm{c}|\mathrm{o})$, $z = 2.04$, $p = .04$.

Discussion

The Exp. II data provide a close replication of Exp. I, supporting the notion that improved recall in re-presentation occurs if and only if some of the letters were originally correct. It seems likely that this effect is due to the effective shortening of the repeated item because of previously remembered letters. The evidence for this hypothesis and its refinement, which postulates that single-letter retention also depends on the number of unremembered letters in the re-presented item, is shown in Tables 2–5. Both item-level analyses reveal that the probability of being correct after repetition depends on the number of missed letters, with a minimum at the probability of being correct either on repeated items with every letter previously incorrect or on new items. Additional and more stable evidence that completely missed items behave as new items is the near identity of letter performance on new items and items completely missed initially.

Evidence for the notion that the single-letter "memory trace" can be represented by an all-or-none process comes both from (*a*) the failure to differentiate repeats completely missed initially from new items, either in letter or item recall, and (*b*) the failure to find a significant intertrial interval effect in Exp. II. If missed letters retained some subthreshold memory trace, performance on repeats should decline with an increasing intertrial interval, just as original recall declines with an increasing retention interval.

EXPERIMENT III

The third experiment was designed as a strong test of the hypotheses suggested by Exp. I and II. In Exp. III, *S* was tested on only one of the three letters originally presented. If a recall error was made, he received a repeat trial on a "related" item, followed by another single-letter recall test. The subset of repeat trials that is of main experimental interest involved a letter test at the same letter position as originally tested. Suppose *S* sees BCD, is tested for the recall of B, and makes an error. There were four groups of *S*s that differed in the nature of the repeated item: (*a*) Group 1 *S*s received a repetition of the same item, BCD; (*b*) Group 2 *S*s had a new letter substituted for B, e.g., J, and received the item JCD for their repeat trial; (*c*) Group 3 *S*s had only the untested letters replaced for their repeats, e.g., BXZ; and (*d*) Group 4 *S*s had both the tested letter and the untested letters replaced, e.g., JXZ. Thus, the four conditions fall into a 2 × 2 design involving the same or different tested letter (e.g., B vs. J) and the same or different context (e.g., CD vs. XZ). If the hypothesis of single-letter all-or-none forgetting, with relearning probability dependent on the context, holds, then the main effect of context ought to be significant, with best performance occurring when the context is the same on the repeat trial as it was on the original trial. If the single-letter forgetting process is all-or-none, fresh letters could be substituted for missed letters without finding a significant main effect of letter.

Method

Subjects and design.—The *S*s were 84 undergraduates from the University of Illinois, whose participation fulfilled a course requirement. A factorial design with the aforementioned two factors was employed. Twenty-one *S*s were randomly assigned to each of the four conditions.

Apparatus.—A Kodak Carousel projector was used to present the slides. The presentation order for one complete trial was: "Ready" signal (2 sec.), CCC item (1 sec.), retention interval (15 sec.), recall interval (7 sec.), and intertrial interval (7 sec.).

Procedure.—Each *S* was run individually. To insure that *S* learned each item, *S* was instructed to read every item aloud during acquisition. During the retention and intertrial intervals, *S* saw a matrix consisting of five rows of random digits with 12 digits in each row and was instructed to read the digits aloud as fast as possible. At the beginning of the recall interval, a slide appeared showing two asterisks and a dash (e.g., **—) and *S* was instructed to say aloud the letter that had originally appeared in the position marked by the dash. Each letter position was cued approximately equally often.

If *S* correctly recalled the cued letter, a new original item appeared on the next trial. However, if *S* missed the requested letter, on the next trial he was exposed to an item related to the one just missed. The four conditions in the experiment differed only in the nature of the related items that were presented following single-letter errors. Basically, a related item could have the same or different letter (LS or LD) and the same or different context (untested) letters (CS or CD). These variables comprised the two factors in the design. Group LS-CS always saw the same item following a single-letter error; Group LS-CD saw an item with the same tested letter but two new context letters; Group LD-CS had only the tested letter changed; and Group LD-CD saw an item with all three letters changed. The *S*s were run until they made eight single-letter errors or until the 48 items were exhausted; in the latter case, they were eliminated from the experiment and replaced by new *S*s.

Following errors on tested letters, the related items were tested four times at the same letter position, interspersed with four times at a different letter position. The data for all of the relearning analyses came from the four trials where each *S* was retested in the same letter position following an error.

Items.—All of the CCC items had a Witmer (1935) association value of 25% (this included the transformed repeat items). None of the items contained the letters M, N, V, or W. There were four sets of items, one for each group of *S*s, devised as follows: First, a core list of 48 items was assembled, along with the selection of a letter position to test in each item. Then, a replacement list of 48 yoked, related items was assembled for each group. These replacement lists corresponded to the group conditions previously mentioned (e.g., for Group LS-CD, the yoked item had the same letter in the tested position but new context letters).

For the first *S* in each group, all of the original items came from the core list and all of the repeat items following errors came from the replacement list. Missed original items and their yoked repeat items were juxtaposed for the next *S* to insure that each item would appear approximately equally often as an original item and as a repeat item. Thus, as the experiment progressed, the current sets of original items and repeat items were each comprised of both core and replacement list items (where core and replacement are defined in terms of the two lists for the first *S*).

Results

Letter-level analysis.—The dependent variable came from repeat trials where the letter to be recalled was in the same serial position as the letter forgotten on the previous trial. There were four such trials for every *S*. The proportion of correct responses on these trials was determined for each *S*; the mean of these proportions for each group is shown in parentheses in Table 6.

TABLE 6

MEAN PROPORTION OF CORRECT RESPONSES ON REPEAT TRIALS WITH RECALL AT SAME SERIAL POSITION AS ON PREVIOUS TRIAL

Context letters	Tested letter		
	Same as previous trial	Different from previous trial	$\bar{X}$
Same as previous trial	**.782** (.762)	**.750** (.738)	**.765** (.750)
Different from previous trial	**.585** (.608)	**.512** (.525)	**.550** (.565)
$\bar{X}$	**.682** (.685)	**.632** (.630)	

Note.—Numbers in boldface are adjusted means; numbers in parentheses are nonadjusted means. Analysis of covariance: $MS_e = .713$; analysis of variance: $MS_e = .811$.

The frequency of correct responses (i.e., proportion correct × 4) for each *S* was entered into a two-way analysis of covariance. The two main effects were tested letter (LS or LD) and nontested context letters (CS or CD). There were two covariates in the analysis, corresponding to two possible sources of confounding in Exp. III. One covariate adjusted for any intergroup differences in inherent recall skill that may have survived the random assignment of *S*s to groups. Notice that all four groups began with the same core list of items. Thus, any differences in the groups' proportions of core list items correct during original trials can be attributed to varying inherent recall skill. This covariate adjusted the dependent variable score of each *S* so that inherent recall skill would not bias the results (the actual entry for this covariate was the arc sine of the proportion of core list items correctly recalled during original trials). The second covariate adjusted for differences in the inherent difficulty of items from the four replacement lists. Since each group had different replacement lists, it was possible that differences on repeat trials could be due to easier replacement lists rather than to the contingency between original and repeat items being investigated. Thus, each *S*'s dependent variable score was adjusted by his proportion of correct responses on those original trials containing items from the replacement list (again, the actual entry into the analysis was the arc sine of the proportion).[2] The correlation between the first covariate and the dependent variable was +.189, the correlation between the second covariate and the dependent variable was +.374, and the multiple correlation between both covariates and the dependent variable was +.378. The mean proportion of correct responses on same-position-cued repeat trials, after covariate adustments, is shown in boldface type in Table 6.

---

[2] For *S*s in Group LS-CS and for the first *S* in each remaining group, their first-covariate entries were also used as their second-covariate entries because there were no distinguishable replacement list items on original trials for these *S*s.

Recall was significantly better when the context remained the same than when it changed, $F\ (1, 78) = 21.22,\ p < .001$. The same-context-vs.-changed-context variable accounted for 20% of the total variance in the dependent variable. The difference between the recall of same and changed tested letters did not even approach statistical significance, $F\ (1, 78) = 1.13$. The Context × Tested Letter interaction was also nonsignificant, $F\ (1, 78) = .19$.

It may be noted that a two-way analysis of variance with no covariance adjustments yielded results that paralleled those from the analysis of covariance. The respective *F*s, each based on 1 and 80 *df*, were 14.11 ($p < .001$), 1.19, and .37.

Phonemic analyses.—Since previous research has shown that errors in STM are often phonemic confusions (Conrad, 1964; Wickelgren, 1966), some improvement might be expected during repetition if the original error response was phonemically related to the correct letter (i.e., "partial retention" of the correct letter).

Each of the 84 *S*s contributed eight original-trial errors to the phonemic analysis. Three classes of consonants were formed on the basis of a common vowel phoneme (phonemicization from Gleason, 1955): Class C_1 consisted of [B, C, D, G, P, T, V, Z], which share an /iy/ phoneme; C_2 consisted of [F, L, M, N, S, X], sharing an /e/ phoneme; and C_3 consisted of [H, J, K], sharing an /ey/ phoneme. Table 7 presents a 3 × 3 confusion matrix whose *ij*th term is the number of times the error response to a letter in Class C_i was in C_j.

TABLE 7

FREQUENCY OF ORIGINAL-LETTER ERRORS AS A FUNCTION OF CORRECT AND OBTAINED VOWEL PHONEMES

Correct vowel phoneme	Obtained vowel phoneme		
	/iy/	/e/	/ey/
/iy/	152 (124.2)	40 (58.6)	91 (97.8)
/e/	68 (86.5)	38 (26.7)	58 (58.1)
/ey/	54 (55.5)	30 (22.3)	23 (24.4)

Note.—Entries without parentheses are observed frequencies; entries in parentheses are expected frequencies.

To test the null hypothesis that vowel phoneme similarity did not account for errors on original items, expected frequencies were computed. Unfortunately, the independence assumption does not imply that the expected frequencies can be obtained directly from the marginal frequencies in Table 7 because, e.g., when the correct letter is H, only J and K could be given as C_3 errors. The expected frequencies were computed by first considering the number of times that a particular letter was tested. Then, the frequencies of errors in the three classes were determined under the assumption that the remaining consonants would be given as errors independently of the phoneme class of the presented letter. These expected frequencies are shown in parentheses in Table 7.

If phonemic confusions did account for errors, then the observed frequencies on the main diagonal of Table 7 (which reflect errors in the same phonemic class as the correct letter) should be higher than their corresponding expected frequencies; the reverse pattern should occur for off-diagonal entries. It is clear that this pattern generally holds, and a chi-square test comparing expected and observed yielded $\chi^2 = 21.8,\ p < .01$.[3] Thus, there is evidence for the hypothesis that letter errors tend to be related phonemically to the correct letter, although the evidence does not reveal a strong relationship.

Even though phonemically related errors occurred in original recall, phonemic similarity played no systematic role in predicting performance on re-presented letters. When *S*'s error was in the same vowel class as the correct response and the same letter was repeated and tested, the proportion correct was .698 $(N = 53)$; when *S*'s error was in a different vowel class, the analogous proportion correct was .661 $(N = 115)$. These two proportions did not differ significantly, $z = .47, p = .64$.

---

[3] The correct degrees of freedom for the test are not easily obtained since a letter-level guessing model was used to compute the expected frequencies. However, the degrees of freedom are obviously between 4 and 8; the reported significance level is for $df = 8$.

Transposition analyses.—Letter errors can be analyzed not only phonemically, but also in terms of whether they come from within the item (transposition) or from outside the item (intrusion). It is known that transpositions account for a large proportion of errors in STM (e.g., Conrad, 1965). Table 8 presents a 3×4 matrix whose *ij*th term is the frequency of times that the requested letter was in Serial Position i and the error response was in Serial Position j $(j = 1, 2, 3)$ or from outside the item $(j = \text{intrusion})$.

Table 8 indicates that 42% of all errors were transpositional. Also, when the requested letter was in the first or third serial position, significantly more adjacent transpositions occurred than remote transpositions, $\chi^2\ (1) = 25.29, p < .001$. Finally, a strong serial position effect is revealed in the row marginals, indicating that fewer errors are made when the requested letter is in the first, then third, and finally, second serial position, $\chi^2\ (2) = 28.67, p < .001$.

Since transpositions account for original errors, perhaps single letters missed through transpositions will show better performance following re-presentation than letters originally missed through intrusions. If so, the recall following re-presentation should be higher when the original error was a transposition then when it was an intrusion. For same-context repeats, the proportion correct (following re-presentation), conditional on the same tested letter and a previous transposition error, was .806 $(N = 36)$; the same conditional proportion with intrusion rather than transposition was .729 $(N = 48)$, z for difference $= .83$, $p = .41$. The analogous proportions when the originally tested letter was replaced by a new letter were .758 $(N = 33)$ for previous transposition and .726 $(N = 51)$ for previous intrusion, z for difference $= .33$, $p = .74$. Thus, in the present experiment (where context letters were not tested), transpositions had a nonsignificant facilitative effect; whatever influence they do have in relearning missed letters does not seem to be potent. Transpositions may have more of an effect on the recall of re-presented *items* than on the recall of re-presented missed *letters*.[4] Additional research is necessary to untangle the relearning benefits of transpositions from the relearning benefits of retained context letters.

TABLE 8

Frequency of Original-Item Errors That Were Transpositions and Intrusions as a Function of Requested Letters's Serial Position

Serial position of requested letter	Error: Position transposed from 1	2	3	Intrusion	Σ
1	—	42	22	109	173
2	60	—	64	161	285
3	23	65	—	126	214

Discussion

Experiment III provides strong evidence for the hypotheses suggested by Exp. I and II. First, a missed letter in a CCC can be replaced by a new letter with no decrement in the recall following re-presentation. Thus, for the task considered, the recall of single letters appears all-or-none. Second, if a missed letter or a substitute is re-presented in the old context, subsequent recall is better than when re-presentation is in a new context, presumably because one or two letters of the old context are not forgotten prior to re-presentation; consequently, the memory load for the repeated item is less.

The phonemic and transposition errors on original trials seem to be in accord with an associationistic theory of memory. However, the failure of these kinds of errors to facilitate relearning suggests that they did not strongly contribute to associative strength. The *S*'s memory may be cluttered with phonemic residuals from partially forgotten letters; hence, it is quite possible that many of these traces will not be of any value as sources of recall strength to build on during re-presentation. Contrariwise, remembering some letters of a multiletter item clearly does benefit the relearning of missed letters. These observations might suggest some sort of modified "slot" model (cf. Neisser, 1967), where the originally presented CCC initially occupies three slots. The *S* tries to consolidate the three letters; however, if a given letter vacates a slot, another letter intruding into that slot or a letter from an adjacent slot is retrieved erroneously. Subsequently, correctly retained letters could provide some sort of address (in the computer sense) that allows the re-presented item to "connect" with its former presentation. This might facilitate the recall following re-presentation in at least two ways: (*a*) The "connected" repetition may cause two letters to converge into a single chunk, thereby reducing the memory load (e.g., through repetition C, A, T may begin as three chunks and converge into the single chunk CAT; cf. Hellyer, 1962); or (*b*) remembering two letters of a trigram may allow more selective attention to be directed at a previously forgotten or new letter (cf. the control processes mentioned in Atkinson & Shiffrin, 1968).

---

[4] Transpositions tended to facilitate relearning at the item level in Exp. I and II (the data were not reported because of small *N*), but only when the original-trial response contained two incorrect letters $(p = .11)$; there was hardly any difference between re-presenting transpositions and intrusions when all three letters were originally incorrect $(p > .50)$.

CONCLUSIONS

There does not seem to be any single simple answer to the decremental vs. all-or-none forgetting question. Whatever answer is provided is bound inextricably in the frame of reference used to view the problem. If one considers the CCC at the level of the entire item, evidence can be found for decremental and multistate forgetting (e.g., the states might be contingent on the number of originally remembered letters). If one views the CCC at the level of the single letter, forgetting seems to occur either completely or not at all. In an effort to discover a residual letter trace, phonemic confusions in single-letter recall were examined. Letter errors did tend to be phonemically similar to the correct letter; however, there was no evidence that these phonemic errors represented residual information that could facilitate the recall following relearning.

The present experiments focused on recall. Another facet of STM is recognition, and failure in recall (even if it is all-or-none) does not necessarily imply failure in recognition (McFadden & Greeno, 1968). For example, the phonemic and transpositional factors accounting for initial recall errors might have facilitated correct recognition.

The formulation that seemed to be most parsimonious in explaining the present data has the single letter as its frame of reference. This explanation consists of two propositions: (*a*) Remembering one or more letters of a multiletter item benefits the recall, following relearning, of a previously forgotten letter; however, (b) subsequent recall is no better than it would have been if a new letter had been substituted for the previously forgotten letter prior to relearning. One goal of future research might be to elaborate such a formulation so that the forgetting of more meaningful material, such as sentences and paragraphs, can be understood.

REFERENCES

Atkinson, R. C., & Crothers, E. J. A comparison of paired-associate learning models having different learning and retention axioms. *Journal of Mathematical Psychology,* 1964, **1**, 285–315.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

Bernbach, H. A. A forgetting model for paired-associate learning. *Journal of Mathematical Psychology,* 1965, **2**, 128–144.

Bower, G. H. A model for response and training variables in paired-associate learning. *Psychological Review,* 1962, **69**, 34–53.

Bregman, A. S. Distribution of practice and between-trials interference. *Canadian Journal of Psychology,* 1967, **21**, 1–14.

Calfee, R. C., & Atkinson, R. C. Paired-associate models and the effect of list length. *Journal of Mathematical Psychology,* 1965, **2**, 255–267.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology,* 1964, **55**, 75–84.

Conrad, R. Order errors in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 161–169.

Gleason, H. A., Jr. An introduction to descriptive *linguistics.* New York: Holt, 1955.

Hellyer, S. Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology,* 1962, **64**, 650.

McFadden, D., & Greeno, J. G. Evidence of different degrees of learning based on different tests of retention. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 452–457.

MELTON, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 1–21.

MURDOCK, B. B., JR. The retention of individual items. *Journal of Experimental Psychology,* 1961, **62**, 618–625.

NEISSER, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology,* 1959, **58**, 193–198.

ROCK, I. The role of repetition in associative learning. *American Journal of Psychology,* 1957, **70**, 186–193.

SNEDECOR, G. W. *Statistical methods.* Ames: Iowa State University Press, 1956.

TULVING, E. Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychological Review,* 1964, **71**, 219–237.

WAUGH, N. C., & NORMAN, D. A. Primary memory. *Psychological Review,* 1965, **72**, 89–104.

WICKELGREN, W. A. Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology,* 1966, **71**, 396–404.

WITMER, L. R. The association value of three-place consonant syllables. *Journal of Genetic Psychology,* 1935, **47**, 337–359.

(Received January 31, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 107–114

CUEING FUNCTION OF FRAGMENTS OF VERBAL ITEMS

IN-MAO LIU[1]

National Taiwan University

Three experiments investigated the cueing function of word and nonword fragments of various sizes and positions. An orderly change in the cueing function was always obtained when the fragment's size in a given position was continuously varied. When a list of items was presented in a constant order, the results suggest a possible connection between serial learning and paired-associate learning. Manipulation of the fragment's position produced a pattern of cue effectiveness resembling a serial position curve. The pattern was more pronounced in cases of real words and long words than in cases of nonwords and short words. The threshold–frequency relationship and the spew hypothesis were deduced as corollaries of the present findings.

Since the work of Mueller and Pilzecker (1900), the method of presenting word fragments on test trials in verbal learning has been widely used. By means of presenting word or nonword fragments of various sizes and positions, the present experiments attempted to study their function as cues for the reinstatement of the whole words or nonwords.

EXPERIMENT I

In the present experiment a list of nonwords was presented in a constant order. From the standpoint of experimental analysis, the manipulation of fragment size, then, offers some clue as to a possible connection between serial learning and paired-associate learning procedures. For the sake of convenience, suppose there are only four words in a list, each consisting of four syllables, $A_iB_iC_iD_i$, where i denotes the ordinal position of the word in the list, $i = 1, 2, 3, 4$. In a constant-order paired-associate procedure, the list of $A_iB_iC_iD_i$ is presented on acquisition trials, and A_iB_i- - (or A_i- - -, or $A_iB_iC_i$-) is presented on test trials. In response to "stimulus words," S has to recall the missing syllables or "response words." Going one step further and presenting the "fragment" - - - - on test trials, it is found that it is a serial learning procedure, for S has to recall each item in its proper position. Of course, fragments of various sizes can be presented to the same S. One such test list is as follows: $A_1B_1C_1$-, - - - -, A_3- - -, A_4B_4- -. When S is instructed properly in advance, in responding to the second stimulus fragment, he has to try to reproduce the second whole word in the acquisition list. By testing each S with a different test list, such as A_1- - -, A_2B_2- -, - - - -, $A_4B_4C_4$-, it is possible to obtain not only data comparable to those of paired-associate studies, but also data comparable to those of serial learning experiments. At a glance, because the list is presented in a constant order, the former data might not represent a typical case, since S can master the task just by learning the response words in a serial order. However, if this happened, there would be an abrupt discontinuity in the data obtained for the paired-associate cases with fragments of varying sizes as stimulus words. In other words, if S concentrates on memorizing the sequence of C_iD_i, he will fail absolutely to recall $B_iC_iD_i$ in response to A_i- - -. Then there will be a discontinuity in recall data, corresponding to the paired-associate case of A_i- - - as stimulus words and the paired-associate case of A_iB_i- - as stimulus words.

When it is assumed that the more parts of the whole a stimulus fragment contains, the greater is the similarity, a question naturally arises as to whether the cueing function of a stimulus fragment depends on its position. Specifically, the problem is whether recall obtained with the stimulus fragments - - - -, A_i- - -, A_iB_i- -, $A_iB_iC_i$- varies in the same way as that obtained with the stimulus fragments containing the final parts, viz., - - - -, - - - D_i, - - C_iD_i, - $B_iC_iD_i$. This problem is also related to the principle of associative symmetry (Asch & Ebenholtz, 1962).

---

[1] The author wishes to express his appreciation to Lin Han and Ai-li Ko for collecting part of the data of Exp. I and to Sin-tsi Lin for collecting part of the data of Exp. II.

Requests for reprints should be sent to In-mao Liu, Department of Psychology, National Taiwan University, Taipei, Taiwan, Republic of China.

Method

Subjects.—The *S*s were 20 male and 20 female undergraduate students of various department majors. They voluntarily participated in the experiment. They had at least 6 yr. of high school English.

Procedure.—The list used in the present study contained eight nonwords as follows: WuHuZuDu, YaSaNaGa, FoKoBoPo, RiTiMiCi, SuYuGuBu, DaMaCaFa, ZiWiPiKi, NoToRoHo. All consonants of a nonword were distinct. Each consonant appeared twice in the list. The first syllables of all nonwords differed from one another, and so did the last syllables.

First, *S* was given a practice task. Three nonwords(MaNa, BeTe, ViWi) appeared one at a time on a Gerbrands memory drum. Each item lasted 1.5 sec., and a 1.5-sec. blank space separated one item from the next. After *S* studied this short list, a fragment of each item was presented according to the order of the list. The *S* was asked to supply the missing syllables as soon as possible. First, *S* was tested with the fragment Ma--, then with ----, and then with --Wi. When the second fragment ---- was presented, he was told to write down the entire second item of the list. He was reminded, therefore, to study not only items, but also their order.

After *S* completed the practice task, he watched the experimental list appear on the memory drum. The eight nonwords appeared one at a time. Each nonword lasted 1.5 sec., and a 1.5-sec. blank space separated one item from the next.

After every four presentations of the list, a test trial was run. For each item, taking WuHuZuDu as an instance, one of the following eight possible ways of testing was used: WuHuZu--, WuHu----, Wu------, --------, --HuZuDu, ----ZuDu, ------Du, --------. On a test trial, since the list contained eight items, each possible type of test appeared only once. Though -------- appeared twice in a test list, it was used once to test forward association and again for backward association. This arrangement was used merely as a statistical convenience. An example of a test list is as follows: WuHu----, YaSaNa--, --------, ------Po, Ri------, --YuGuBu, ----CaFa, --------. If *S* did not start to write down the missing syllables for each stimulus fragment within 4 sec. after its presentation, the next stimulus fragment was presented.

As was shown in the example of WuHuZuDu, since there were eight possible types of test for each item, eight test lists were generated. Every possible type of test for each item appeared once in one of the eight test lists.

For each *S*, the acquisition list was presented 20 times. Hence, five test trials were administered for each *S*. On each test trial, a different test list was used. Five distinct test lists for each *S* out of the eight test lists were systematically varied from *S* to *S*.

The acquisition list was arranged in two orders. For each acquisition list of one order, there were eight test lists of that order. One order of the acquisition list was used for half of the *S*s, and the other order was used for the other half of the *S*s.

Results and Discussion

Table 1 shows the proportion of *S*s giving correct recall to each type of stimulus fragment. Obviously, both increases in fragment size and training facilitated recall. Recall was slightly better when the stimulus fragment was in the initial position than in the final position. An analysis of variance (40 *S*s × 2 Fragment Positions × 4 Fragment Sizes × 5 Degrees of Training) showed that fragment position was significant, $F\ (1, 39) = 5.60$, $p < .05$. Both fragment size in terms of number of syllables and amount of training were significant, $F\ (3, 117) = 86.35$, $p < .001$; and $F\ (4, 156) = 58.41$, $p < .001$, respectively. All the interaction effects were not significant: $F\ (3, 117) = 2.27$, for Fragment Position × Fragment Size; $F\ (4, 156) < 1$, for Fragment Position × Amount of Training; $F\ (12, 468) = 1.24$, for Fragment Size × Amount of Training; and $F\ (12, 468) = 1.41$, for Fragment Position × Fragment Size × Amount of Training.

Although, in general, the initial fragment served as a better cue than the final fragment, only the difference between the initial and final fragments containing two syllables was significant, $t\ (39) = 2.89$, $p < .01$. The same difference but of smaller magnitude was observed between the initial and final fragments containing one syllable, but it was not significant, $t\ (39) = 1.84$.

The differences in the cue function of the initial fragments containing any two different numbers of syllables were all significant: $t\ (117) = 3.33$, $p < .01$, for zero and one syllables; $t\ (117) = 3.61$, $p < .001$, for one and two syllables; $t\ (117) = 4.17$, $p < .001$, for two and three syllables. For the final

TABLE 1

PROPORTION OF *S*s GIVING CORRECT RECALL TO INITIAL (I) AND FINAL (F) FRAGMENTS

Syllables in fragment	Test trial											
	1		2		3		4		5		Total	
	I	F	I	F	I	F	I	F	I	F	I	F
0	.00	.03	.05	.20	.23	.13	.28	.35	.58	.43	.23	.24
1	.08	.10	.30	.05	.33	.20	.45	.45	.58	.58	.35	.28
2	.23	.03	.35	.20	.43	.48	.63	.53	.75	.60	.48	.37
3	.38	.35	.55	.60	.63	.68	.73	.73	.85	.73	.63	.62
Total	.17	.13	.31	.26	.40	.37	.52	.51	.69	.59		

fragments, the corresponding t's were $t\ (117) = 1.11$ (not significant); $t\ (117) = 2.50$, $p < .05$; and $t\ (117) = 6.94$, $p < .001$.

An orderly gradient of recall along each column of Table 1 suggests that recall to the zero-syllable fragment, which is what might be obtained in a serial learning experiment, may be considered a special case of the fragment containing one, two, or three syllables. It further suggests that though the list was presented in a constant order, *S*s did not concentrate on some particular response syllables to be later tested.

EXPERIMENT II

In Exp. I, it was found that there exists a sort of asymmetry between the cueing power of initial and final fragments. However, these results could be weakened by and confounded with the effect of presenting the list in a constant order. Some investigators (Carluccio & Crowder, 1966; Martin & Saltz, 1963; Newman & Forsyth, 1965; Newman & Saltz, 1962) found that the function of the serial position cues in paired-associate learning is small and negligible. However, other experiments (e.g., Brown, 1968) have pointed to the use of serial position cues in constant-order paired-associate learning.

With respect to the fragment position effect obtained in Exp. I, some particular order of parts of an item might favor recall to one fragment position more than to another. This point is connected with the concept of availability (Asch & Ebenholtz, 1962; Horowitz, Brown, & Weissbluth, 1964).

There are other fragment positions other than the initial and final ones. In considering the nonword WuHuZuDu, e.g., a fragment containing one syllable may be --Hu---- or ----Zu--; a fragment containing two syllables may be --HuZu--, Wu--Zu--, Wu----Du, or --Hu--Du; and a fragment containing three syllables may be Wu--ZuDu or WuHu--Du. These fragments are referred to as in the random position. If the cue function of a fragment depends on its position, the random position also has to be taken into consideration.

In view of the three problems raised above, Exp. II was designed. First, the order of items in a list was varied from one trial to the next. Hence, the fragment containing zero syllables was not considered. Second, four lists were constructed so that each syllable of an item would be in the first, second, third, or last position in one of the four lists. Third, the random-position fragment was also presented on test trials.

Method

Subjects and materials.—The *S*s were 24 freshmen (8 male, 16 female) from an introductory psychology class.

There were four lists of nonwords. Six *S*s were assigned to each list according to the order of their appearance. The first list contained the following items, which were similar to those of Exp. I: WuHuZuDu, YaSaNaGa, FoKoBoPo, RiZiTiCi, SuYuGuBu, DaMaCaFa, MiWiPiKi, NoToRoHo, VeLeQeJe. If a consonant appeared in one position of an item, it did not appear in the same position for any other item. The second list was obtained from the first by reversing the order of consonants of each item, e.g., WuHuZuDu of the first list became DuZuHuWu in the second list. The third list was obtained from the first by exchanging the first two syllables with the last two syllables for each item, e.g., WuHuZuDu of the first list became ZuDuWuHu in the third list. The fourth list was obtained from the first by exchanging the first with the second and the third with the fourth syllables of each item. Thus in the four lists, each syllable of an item appeared in each position once.

Procedure.—After a practice task, one of the lists was presented to each S 12 times in the same way as in Exp. I. The order of items in the list varied from one presentation to the next. After every four presentations of the list, a test trial was run. On each test trial, 27 fragments (9 fragments each of a nonword × 3 positions) of nonwords consisting of one, two, or three syllables were presented to S one at a time. Each nonword was tested by three kinds of fragments (each containing the same number of syllables) based on their position: initial, random, or final. For example, the stimulus fragment of two syllables for testing the recall of WuHuZuDu was the initial fragment WuHu----, one of the random fragments --HuZu--, Wu--Zu--, Wu----Du, and --Hu--Du, or the final fragment ----ZuDu.

This testing procedure was similar to that of Horowitz, White, and Atwood (1968). A preliminary observation indicated that S, tested first on the initial fragment of a nonword, e.g., did not necessarily give a better recall to the final or random fragment of the same nonword presented on the second or third test. In other words, the pattern for the initial test was not inconsistent with that for the repeated measures. Moreover, the order of presenting three types of fragments of the same nonword was varied systematically from one test trial to the next within S and from S to S.

Nine fragments used for testing recall of three nonwords contained one syllable, nine used for testing recall of another three nonwords contained two syllables, and nine used for testing recall of still another three nonwords contained three syllables. In this way, nine fragments were used to test recall for the initial position, nine to test recall for the random position, and nine to test recall for the final position. Which three fragments for testing recall of a nonword contained one, two, or three syllables was systematically varied from one test trial to the next within S and from one S to another.

Results and Discussion

Table 2 shows the mean number of items that S recalled correctly on the three test trials, taken together for each type of stimulus fragment containing one, two, or three syllables. An analysis of variance indicated that the number of syllables contained in the stimulus fragment was a significant source of variation, $F\ (2, 46) = 114.14$, $p < .001$. Recall was best when the initial fragment served as the cue and poorest when the final fragment served as the cue. The fragment's position in the nonword had a significant effect on recall, $F\ (2, 46) = 4.50$, $p < .05$. The interaction of the fragment's position with the number of syllables contained in the stimulus fragment was not significant, $F < 1$.

TABLE 2

MEAN NUMBER OF ITEMS CORRECTLY RECALLED FOR NONWORDS

Syllables in fragment	Fragment given as cue					
	Initial		Random		Final	
	M	*SD*	*M*	*SD*	*M*	*SD*
1	.96	1.43	.83	1.13	.67	.87
2	2.67	1.73	2.29	1.90	2.08	1.69
3	4.08	1.98	3.67	1.49	3.42	1.84

Note.—Maximum score = 9.

When pairwise comparisons of the fragments' positions were made, the initial position was significantly different from the final position, $t\ (46) = 3.02$, $p < .05$. The differences in recall for the initial vs. the random fragment and for the random vs. the final fragment were, however, not significant, $t\ (46) = 1.80$; and $t\ (46) = 1.22$, respectively. Since the initial fragment aided recall significantly more than the final fragment, a further analysis was made. The initial fragment containing three syllables served as a better cue than the final fragment containing three syllables, $t\ (46) = 2.20$, $p < .05$. The difference for the initial and final fragments containing two syllables was large but not significant, $t\ (46) = 1.97$. A smaller difference in the same direction was obtained for the initial and final fragments containing one syllable.

Table 2 further shows that the effect of the number of syllables contained in the stimulus fragment was greater than that of the fragment's position. More specifically, the three-syllable final fragment (poorest position) still aided recall more than the two-syllable initial fragment (best position), $t\ (23) = 1.96$, $p < .05$ (one-tailed test). The two-syllable final fragment (poorest position) aided recall significantly more than the one-syllable initial fragment (best position), $t\ (23) = 4.15,\ p < .001$. Obviously, one-, two- and three-syllable fragments, irrespective of their position, still produced a clear gradient of effect.

In general, Exp. II supported the findings of Exp. I. When the serial position cues were not available, *S* had to depend on the fragment alone for the cueing function. Therefore, the number of syllables constituting the size of the fragment had a more decisive effect than the factor of the fragment's position.

EXPERIMENT III

In Exp. II, although the random fragment containing two or three syllables was confounded with the initial and final positions, that containing just one syllable was certainly a middle fragment. In view of the finding of Horowitz et al. (1968) that the middle fragment was the poorest in serving as the cue, there are two possibilities to account for this discrepancy in the findings.

First, the discrepancy may be due to word vs. nonword differences. When a list consists of nonwords, as in Exp. II, *S* has to study each part of an item equally anew. On the other hand, with words, as in the Horowitz et al. (1968) experiment, differential cue properties associated with each part of a word might have been established through perceptual learning. As they have pointed out, this interpretation is not sufficient since it does not explain why kindergarten children, before they learn to read, show the adult pattern in judging the similarity of nonsense syllables (Marchbanks & Levin, 1965).

Second, the discrepancy in the findings may be accounted for in terms of word length differences. In the experiment of Horowitz et al. (1968), words of six and nine letters were used. Although nonwords of eight letters were used in Exp. II, since vowels within each item were the same and were used to aid pronunciation, actually there were only four units or "chunks" (Miller, 1956). This inference based on the difference in lengths of items is reasonable if learning of a list of words, each of a certain length, is regarded as a kind of serial learning taking place simultaneously for each word. Then, differences in the cueing power of initial, middle, and final fragments of each word correspond merely to a serial position curve. If the length of words in the list is increased, the middle portion of these miniature serial position curves would be expected logically to be elevated.

In this experiment, then, the length of the items in the list was manipulated; in addition, real words instead of nonwords were used.

Method

Subjects.—The *S*s were 24 female seniors in the English department. They studied at least six years of English in high school and more than three years of English at the university. The frequency count of English words undoubtedly applies to these students. They voluntarily participated in the experiment.

Materials.—Four lists of words were used; two lists contained nine-letter words, and two contained four-letter words. The two lists of nine-letter words each contained eight words: EXCELLENT, AFTERWARD, DANGEROUS, STRUCTURE, BEAUTIFUL, FOLLOWING, INSTANTLY, PHYSICIAN for one list, and DISCOVERY, UNIVERSAL, CHARACTER, VEGETABLE, PRESIDENT, ESTABLISH, ACCORDING, RELIGIOUS for the other. The two lists of four-letter words each contained nine words: ARMY, FLOW, STEP, INCH, HAVE, CORN, DESK, WIND, QUIT for one list; and COLD, HEAT, FIVE, DARK, THUS, UPON, OBEY, SUCH, INTO for the other.

The words were all high-frequency words in the Thorndike-Lorge (1944) word count; their frequencies were 35 or more. As to the lists of nine-letter words, the first, fifth (middle), and last letters of the words in a list differed from one another. For the rest of the positions, there were, at most, two identical letters for two words in the same list. With respect to the four-letter words in the same list, there was no identical letter appearing in the same position.

Procedure.—After *S* completed a practice task (a list of CAT, TOY, PEN), she watched one experimental list appear on the memory drum in the same way as in Exp. I. After *S* studied the entire list, she was shown a fragment of one word she had seen and was asked to supply the entire word. The *S* was allowed only 4 sec. for responding. When *S* responded (or when 4 sec. had elapsed), the next fragment was presented.

Each *S* studied one of the two lists of nine-letter words and one of the two lists of four-letter words. List assignment and order of assignment were systematically varied from *S* to *S*. Eight *S*s were tested with one-letter fragments, another eight *S*s were tested with two-letter fragments, and

TABLE 3

Mean Number of Items Correctly Recalled for Nine-Letter Words

Letters in fragment[a]	Fragment given as cue									
	Initial		Middle						Final	
			Clustered		Random		Combined			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	4.88	1.36	2.13	.84	2.63	.74	2.38	.52	3.88	1.24
2	7.50	.53	4.13	1.55	3.88	1.13	4.00	1.04	5.00	.93
3	7.50	.53	6.25	1.75	4.00	.93	5.13	1.09	6.75	.46

Note.—Maximum score = 8.
[a] $N = 8$.

still another eight *S*s were tested with three-letter fragments.

The fragment in the case of a list of nine-letter words was always the initial part of the word, the clustered-middle part, the random-middle part, or the final part. For example, the one-letter stimulus fragment for testing the recall of DISCOVERY was "D--------," "----O----," "--S------," or "--------Y." In the category of the one-letter, random-middle part were also "-----V---," "-------R-," etc. The fragment "------E--" was excluded since both DISCOVERY and PRESIDENT of the same list have the same letter in the seventh position. Likewise, the two letter stimulus fragment was "DI-------," "---CO----" (also in the same class was "----OV---"), "---C---R-" (also in the same class were "-I---V---," etc.), or "-------RY." The three-letter fragment was constructed in an analogous way. Instances of the three-letter, random-middle fragment were "--S--V-R-," etc., as long as the letters of the first, middle, and last positions were not included and not all three letters were clustered together.

The fragment in the case of a list of four-letter words was always the initial part of the word, the middle part, or the final part. For testing the recall of STEP, e.g., the one-letter stimulus fragment was "S---," one of "-T--" and "--E-," or "---P"; the two-letter stimulus fragment was "ST--," "-TE-," or "--EP"; and the three-letter stimulus fragment was "STE-," one of "S-EP" and "ST-P," or "-TEP." Of course, the three-letter, middle-part fragment in the case of four-letter words was actually the extremity-part fragment.

Each word was tested by all four types of fragments in the case of a list of nine-letter words and by all three types of fragments in the case of a list of four-letter words, so that the test lists contained 32 fragments and 27 fragments, respectively.

For some fragments' positions, as mentioned previously, there were several instances belonging to the same fragment's position. In these cases, such instances were systematically varied from one word to another in the same list. There were eight orders of presenting the stimulus fragments containing a particular number of letters. Each *S* was tested with one order.

Results and Discussion

The mean number of items that *S* recalled correctly to each position of stimulus fragment containing a given number of letters is shown in Tables 3 and 4.

With respect to nine-letter words, the cueing power of the clustered-middle fragment and that of the random-middle fragment were not very different when fragments contained only one or two letters. However, when the fragment's size was increased to three letters, the clustered-middle fragment served better as the cue than the random-middle fragment. An analysis of variance applied to the data obtained for these types of fragments showed that the Fragment Size × Fragment Position interaction was significant, $F\ (2, 21) = 6.41,\ p < .01$. The better recall for the three-letter, clustered-middle

TABLE 4

Mean Number of Items Correctly Recalled for Four-Letter Words

Letters in fragment[a]	Fragment given as cue					
	Initial		Random		Final	
	M	*SD*	*M*	*SD*	*M*	*SD*
1	6.38	2.13	4.13	2.03	4.88	2.03
2	8.50	.93	7.00	1.69	8.13	.83
3	8.88	.13	9.00	.00	8.88	.13

Note.—Maximum score = 9.
[a] $N = 8$.

fragment compared with the three-letter, random-middle fragment was significant, $t\ (21) = 4.02, p < .001$. The overall difference between the cueing power of these two types of fragments was, however, not significant, $F\ (1, 21) = 4.24$.

With the cluster-middle and random-middle fragments combined, as shown in Table 3, an analysis of variance applied to all data of nine-letter words indicated that the fragment's position was a significant source of variation, $F\ (2, 42) = 75.30, p < .001$. The fragment's size had a significant effect, $F\ (2, 21) = 36.48, p < .001$. The interaction of these two effects was also significant, $F\ (4, 44) = 2.98, p < .05$. As is clear from Table 3, this interaction was mainly due to the ceiling effect of the cueing function for the initial fragment when the fragment's size was more than two letters.

Table 3 further shows that the difference in the cueing function of the initial vs. the combined-middle fragment was larger than that of the final vs. the combined-middle fragment. Even the latter difference of smaller magnitude was significant, $t\ (42) = 6.06, p < .001$. Comparing the effect of fragment's size across fragment's position, it is found that the three-letter, combined-middle fragment was inferior to the two-letter initial fragment in serving as the cue, $t\ (14) = 5.53, p < .001$. Though not significant, $t\ (14) = 1.45$, the mean recall to the one-letter initial fragment was larger than that to the two-letter, combined-middle fragment. Thus, the fragment's position was more critical than the fragment's size in this case.

With respect to four-letter words, referring to Table 4, the slight advantage in recall for the middle fragment containing three letters, or at least the lack of a difference, apparently occurred because it always contained the initial and the final letters. An analysis of variance showed that fragment position, fragment size, and their interaction were all significant, $F\ (2, 44) = 17.55, p < .001$; $F\ (2, 22) = 25.87, p < .001$; and $F\ (4, 44) = 13.28, p < .001$, respectively. The difference in recall for the initial vs. the final fragment was significant, $t\ (44) = 2.16, p < .05$. Although the cueing function was, in general, better for the final fragment than for the middle fragment, the difference was not significant when the confounded data obtained for the three-letter fragment were excluded from consideration, $F\ (1, 14) = 2.97$.

When the effect of fragment size is examined in Table 4 across fragment position, orderly gradients of recall based only on the fragment's size irrespective of the fragment's position can still be established.

The factor of the word vs. nonword difference as well as that of word length was found to have effects on the cueing function of the fragment's position. For nine-letter words of great length, the cueing function of parts was more heterogeneous: the middle part was poorer in serving as the cue not only in comparison with the initial, but also with the final part. When those middle or random fragments confounded with the initial and final positions were excluded from consideration, this bow-shaped cueing function was found even with four-letter words, but not with four-chunk nonwords of Exp. II.

GENERAL DISCUSSION

The finding that a smaller fragment is capable of evoking the entire word association as the number of trials increases has the following implication: The visual recognition thresholds of words are inversely correlated with the relative frequencies with which words are used (Howes & Solomon, 1951; Solomon & Postman, 1952). This implication holds since a smaller fragment of a word can be recognized in a shorter time interval. As an example, suppose that WuHuZuDu is a high-frequency word and YaSaNaGa a low-frequency word, so that Wu------ alone can make *S* recall the whole, while YaSaNa-- is necessary for *S* to think of the whole. Now, with a 50-msec. tachistoscopic exposure of WuHuZuDu, assume that *S* can identify only one syllable, Wu. Then, *S* can *see* the word at this exposure time. On the other hand, unless the exposure time exceeds, e.g., 70 msec., at which *S* can see YaSaNa--, he cannot recognize the word.

The present finding also implies the spew hypothesis (Underwood & Schulz, 1960), which states that the order of emission of verbal units is directly related to their frequency. Under fairly vague stimulation conditions (e.g., free association or situations where guessing is encouraged), a smaller fragment of word can be contained in a wider range of stimulus patterns. The temporal order of emission of verbal units now follows from the facts that a smaller fragment of a word evokes the association earlier than a larger fragment of a word and that a smaller fragment of a word increasingly aids recall as a function of frequency. As an illustration, suppose that a person has only a vocabulary of common, commit, and comedy and that common is a high-frequency word, while the rest are low-frequency words. Suppose further that co---- enables him to think of *common,* while without COMMI-, he cannot think of *commit.* Then, whenever a vague stimulation or stimulus pattern contains co---- and he tries hard to think of words, the word *common* will be emitted. Now, stimulus patterns containing COMMI- necessarily contain co----. The converse is not true. In other words, the set of stimulus patterns each containing the smaller fragment is not smaller than the set of stimulus patterns each containing the larger fragment. Therefore, first, because of the differential conditions of availability (larger vs. smaller set), *common* will be emitted earlier than *commit.* Secondly, co---- generally will be noticed earlier than the five letters COMMI-, resulting in the earlier emission of *common* than *commit.*

REFERENCES

ASCH, S. E., & EBENHOLTZ, S. M. The principle of associative symmetry. *Proceedings of the American Philosophical Society,* 1962, **106**, 135–163.

BROWN, S. C. Role of stimulus-term and serial-position cues in constant-order paired-associate learning. *Journal of Experimental Psychology,* 1968, 77, 269–274.

CARLUCCIO, L., & CROWDER, R. G. Constant order of pairs in the presentation and testing of paired associates. *Journal of Experimental Psychology,* 1966, 72, 614–616.

HOROWITZ, L. M., BROWN, Z. M., & WEISSBLUTH, S. Availability and the direction of associations. *Journal of Experimental Psychology,* 1964, **68**, 541–549.

HOROWITZ, L. M., WHITE, M. A., & ATWOOD, D. A. Word fragments as aids to recall: The organization of a word. *Journal of Experimental Psychology,* 1968, **76**, 219–226.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *Journal of Experimental Psychology,* 1951, **41**, 401–410.

MARCHBANKS, G., & LEVIN, H. Cues by which children recognize words. *Journal of Educational Psychology,* 1965, **56**, 57–61.

MARTIN, C. J., & SALTZ, E. Serial versus random presentation of paired associates. *Journal of Experimental Psychology,* 1963, **65**, 609–615.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 1956, **63**, 81–97.

MUELLER, G. E., & PILZECKER, A. Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift für Psychologie,* 1900, Ergbd., **1**, 1–288.

NEWMAN, S. E., & FORSYTH, G. A. Isolation effects when paired associates are presented serially. *Journal of Experimental Psychology,* 1965, **70**, 334–335.

NEWMAN, S. E., & SALTZ, E. Serial position as a cue in learning. *American Journal of Psychology,* 1962, **75**, 102–108.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology,* 1952, **43**, 195–201.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received January 31, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 115–120

PROBABILITY LEARNING IN THE CORRECTION T MAZE UNDER NONCONTINGENT REINFORCEMENT SCHEDULES[1]

JANET ROBBINS[2]

Brooklyn College of the City University of New York

Two experiments were conducted to investigate probability learning in the correction T maze using a 24-hr. intertrial interval. In Exp. I, *S*s retraced the entire length of the "incorrect" arm to gain access to the "correct" or reinforced side. In Exp. II, *S*s were restricted to the "choice point" area until a "correct" response was made. Most *S*s in Exp. I tended to go 100% of the time, at asymptote, to one or the other side of the maze, while approximately 60% did so in Exp. II. Theoretical analyses revealed that the Bush and Wilson "secondary reinforcement" model adequately described the behavior in Exp. I, although shortcomings were noted in Exp. II.

Investigators studying probability learning under noncontingent reinforcement schedules in rats in the correction T maze have employed a variety of procedures. When *S* is allowed partially or completely to enter and retrace the incorrect arm, some *E*s have reported "probability matching" at asymptote (Carterette, 1957; Parducci & Polt, 1958), while others have reported "overshooting" of the matching value (Wright, 1967). Matching has also been found with a choice point correction procedure (Brunswik, 1939; Estes, 1957a). Finally, Bitterman and his colleagues (Bitterman, 1965; Bitterman, Wodinsky, & Candland, 1958), using a variety of procedures with rats, have generally found "absorption" or 100% responding. In an attempt to clarify these discrepancies, groups of rats were given probability learning under the conventional procedure of allowing *S* to enter the entire length of the incorrect arm in Exp. I, and under a choice point correction procedure in Exp. II.

Another goal of the present research was to evaluate a class of linear learning models (Bush & Mosteller, 1955; Estes, 1950, 1957b). In all of the models it is assumed that directly reinforced trials result in an increment to the probability of responding to the side chosen. However, correction trials may be assumed to have the following effects on the probability of responding to the reinforced side in contrast to the effect of a directly reinforced trial: (*a*) result in an equal increment (equal θ model of Bush & Mosteller, 1955, and of Estes, 1957b), or (*b*) result in a smaller increment (unequal or 2θ model of Bush & Mosteller, 1955), or (*c*) have no effect, i.e., an "identity operator" (Bush & Mosteller, 1955). Finally, a correction trial may result in an increment to the side chosen, although less than that due to a directly reinforced trial ("secondary reinforcement" model of Bush & Wilson, 1956).

The equal θ model generates an asymptotic prediction of probability matching, while the unequal θ model predicts overshooting. Both the Bush-Wilson and the identity operator model predict absorption either to the high or to the low proportion side at asymptote, the distribution being a function of the rate of learning and the reward proportion.

---

[1] This report is based on a dissertation submitted to the Graduate Faculty in Psychology, Brooklyn College of the City University of New York, in partial fulfillment of the requirements for the PhD degree. This research was supported, in part, by Grant MH-13955 from the National Institute of Mental Health, a grant from the Sigma Xi Research Fund, and a grant from the City University of New York. The author held a National Institute of Mental Health predoctoral fellowship during the course of this investigation. The author is deeply indebted to her sponsor, Solomon Weinstock, for his advice and encouragement. Finally, special thanks are due to the staff at the Brooklyn College Computer Center, especially to Howard Givner and Martin Cohen, who wrote the programs necessary for the analysis of the present data.

[2] Request for reprints should be sent to Janet Robbins, c/o Donald Robbins, Rockefeller University, New York, New York 10021.

EXPERIMENT I

The conventional correction procedure of allowing S to retrace the entire length of the incorrect arm in order to gain access to the "correct" or reinforced side was used. Two groups of rats were given one trial per day under a 60:40 or 70:30 noncontingent reinforcement schedule.

Method

Subjects.—Sixty female albino rats, 100–120 days old at the start of the experiment, were obtained from Albino Farms, Little Silver, New Jersey.

Apparatus.—An enclosed T maze painted flat gray and 5 in. high and 5 in. wide throughout was used. The start box, stem, and cross piece (including the choice area) were 9, 24, and 41 in. long, respectively. An aluminum guillotine door separated the start box and the stem, while aluminum sliding doors, mounted 18 in. from the "end wall" of the goal box, separated each of the arms from the stem. The end walls were 8 in. long with a 3 × 5 in. recessed drinking area in the side walls closer to the start box. A drinking tube similar to those used in the home cages, was mounted on the wall of the recessed drinking area and an aluminum slide could be inserted to prevent access to the tube. The drinking area was not visible from the choice point. The apparatus was covered with hinge-mounted clear Plexiglas. Illumination was provided by one 100-w. bulb centered over the choice point and approximately 6 ft. above the apparatus. An air conditioner which was on continuously provided a uniform masking noise.

Procedure.—Two weeks before the beginning of the experiment, Ss were habituated to a 20-hr. water deprivation schedule and gentled. The Ss were then randomly assigned to groups.

Thirty Ss were given reinforcement on .60 of the trials on one side and .40 on the other side of the maze (Group 60). Another group of 30 Ss were reinforced according to a 70:30 reinforcement schedule on one or the other side of the maze (Group 70). Half of the Ss in each group were given the larger proportion of reinforcement on the right side and the other half on the left side of the maze. Reinforcement for each S was randomly assigned by blocks of 20 trials, with the restriction that each block have reinforcements in the ratio 12:8 for Group 60 and 14:6 for Group 70 and that on each trial reinforcement was available on one or the other side of the maze. Reinforcement was 10 sec. of water consumption, measured on a Standard Electric timer. The Ss were run in a fixed order at one trial per day under a correction procedure in which S was allowed to retrace the entire length of the incorrect arm in order to enter the correct arm. Four reinforced forced trials were given each S, in either a RLLR or a LRRL order, followed by training with correction under the appropriate reinforcement schedule.

Two seconds after S was placed in the start box, the door was raised. A response was defined as the first penetration of 10 in. into either arm of the maze. When S entered the side where reinforcement was present, the sliding door separating the arm from the stem was closed. After completion of 10 sec. of drinking, S was returned to its home cage.

The Ss were given access to water in their home cages for 90 min., approximately 30 min. after the last S was run. Deprivation was thus maintained at 20–22 hr.

Training was continued until Ss appeared to be responding at asymptote and then were given 20 additional trials. As a result, Ss were given 80 trials. These trials were then followed by another block of 20 trials (Trials 81–100) on which reinforcement was available on only one side of the maze, "reversal" training. Reinforcement was placed on the side to which S responded least frequently during the asymptotic block. Therefore, all Ss were switched to a 0:100 reinforcement schedule for the last 20 trials. One S died during the course of the study, leaving 30 and 29 Ss in Groups 60 and 70, respectively.

The side where the larger proportion of reinforcement was available for each S was defined to be the A_1 side for original training. During "reversal," the "preferred" side during the asymptotic block was defined as the A_1 side.

Parameter estimation.—Using all trials, maximum likelihood parameter estimates were made for both Groups 60 and 70 for the equal θ, unequal or 2θ, identity operator (ID), and the Bush and Wilson (BW) models. Using an assumed value of .5 for the initial response probability and the individual reinforcement schedules, one stat rat was "generated" corresponding to each real rat. The two parameters for the Estes N Pattern Model (PM) were estimated using the asymptotic conditional response probabilities suggested by Estes (1959). The value of N was taken to be the next highest integer and the estimate of c was revised accordingly. Stat rats were then generated. Finally, in accordance with the plan decided on in advance of the study, the mean A_1 proportion over Trials 62–81 was taken as an estimate of the asymptotic response probability.

Results

The mean proportion of A_1s during the last 20 trials was .72 and .75 for Groups 60 and 70, respectively. Statistical analyses revealed no significant differences between groups nor any significant difference between each group and its respective reinforcement proportion, $t < 1$. In Table 1 is presented the distribution of the proportion of A_1s during the asymptotic block for Groups 60 and

TABLE 1

ASYMPTOTIC MEAN PROPORTION OF A_1S ($\bar{P}_\infty(A_1)$) AND DISTRIBUTION OF THE PROPORTION OF A_1S DURING THE LAST 20 TRIALS IN EXP. I FOR REAL AND STAT RATS

Item	Frequencies of A_1 proportions				
	$\bar{P}_\infty(A_1)$	.8–1.0	.5–.79	.21–.49	0–.2
Group 60					
Real	.72	19	3	2	6
Equal θ	.57	0	20	10	0
2θ	.81	18	11	1	0
ID	.85	24	2	2	2
BW	.70	19	2	1	8
PM	.59	6	14	7	3
Group 70					
Real	.75	19	5	1	4
Equal θ	.74	3	26	0	0
2θ	.95	27	2	0	0
ID	.96	27	1	1	0
BW	.78	20	4	0	5
PM	.70	8	14	7	0

TABLE 2

PARAMETER ESTIMATES FOR EXP I

Item	Group	
	60	70
Equal θ		
θ	.057	.076
2θ		
θ_1	.160	.200
θ_2	.005	.008
ID		
θ	.190	.150
BW		
θ_1	.200	.150
θ_2	.090	.080
PM		
C	.17	.14
N	2	2

70. These data reveal that most *S*s were absorbed at asymptote with 6 and 4 *S*s from Groups 60 and 70, respectively, absorbed or almost absorbed on the A_2 side.

Table 2 presents the parameter estimates for the various models, and in Table 1 are the distributions of the proportion of A_1s during the last 20 trials generated under the five models being considered. Kolmogorov-Smirnov tests were performed comparing the cumulative frequency distribution of real to stat rats for each model and revealed significant differences in all cases, $K_D \geq 11$, $N = 29$ or 30, $p < .05$, except for the BW model comparisons, $K_D < 4$, $N = 29$ or 30, $p > .05$.

Only the BW model appeared to account for the absorption found at asymptote. As a result, 41 statistics, used by Weinstock, North, Brody, and LoGuidice (1965), were used to explore the fine detail of the data and, in general, revealed fairly close correspondence between real and BW stat rats (see Robbins, 1969, for details).

In "reversal" training, more than 60% of the *S*s made at least 90% of their responses to the absorbed side, thus displaying a considerable amount of resistance to change. The BW stat rats were also "run" for these additional 19 trials, with the real rat asymptotic probabilities as the initial value for these trials. Kolmogorov-Smirnov tests were performed on the cumulative frequency distribution of responses during these 19 trials and revealed no significant differences for both groups, $K_D < 5$, $N = 29$ or 30, $p > .05$.

Discussion

Theoretical analyses revealed that only the BW model adequately accounted for the absorption found at asymptote, the detailed structure of the data, and the resistance to change displayed when the reinforcement schedule was changed. In the BW model, it is assumed that increments occur to the initial side chosen, although larger increments occur if reinforcement is given in that arm of the maze. The application of the model can be justified psychologically in terms of delay of reinforcement. In particular, *S*s appeared to run as rapidly as possible to the other side of the T maze after an "incorrect" response and seemed not to be responding to the choice point stimuli. Correction trials, it may be suggested, therefore differ from immediate reinforcement trials only in that they lengthen the chain of responses before reinforcement or add delay of reinforcement. Furthermore, on a correction trial, except for the very early trials, there was only one turning response, and *S*s typically viewed the choice point stimuli from the stem only at the time of initial choice. Under these conditions, since *S* makes only one turning response in the presence of choice point stimulation, the probability of that turning response may be assumed to be increased when *S* eventually gets reinforced. The longer response chain and therefore the longer delay of reinforcement may also be used to account for the smaller increment resulting from correction trials as opposed to directly reinforced trials.

EXPERIMENT II

In Exp. II, the depth of penetration allowed on correction trials was varied. Two groups of rats were given correction at the choice point, while two others were given correction after a partial or a complete entry into the incorrect arm.

Method

Subjects.—One hundred and two female albino rats, 100–120 days old at the start of the experiment, were obtained from Albino Farms, Little Silver, New Jersey.

Apparatus.—Three enclosed T mazes similar to that of Exp. I were used. The major difference was that the cross piece, including the choice area, was 59 in. long for the choice point groups (Group C), 79 in. long for the group receiving correction 11 in. up the arm (Group I), and 101 in. long for the group allowed to retrace the entire length of the incorrect arm (Group L). The aluminum doors of the arms were mounted 26 in. from the end wall of each goal box, so that the distance from the door in the arms to the choice area was 1, 11, and 22 in. for Groups C, I, and L, respectively. The doors to the arms were mounted in a V-shaped groove which was approximately $\frac{3}{8}$ in. at the top. This permitted the door to be placed in an upright position at the beginning of the trial. The choice response required of *S*s was to push the door $\frac{1}{4}$ in. or more in a horizontal direction as measured at the top of the groove.

Procedure.—The habituation to water deprivation and gentling procedures were identical to those in Exp. I. The *S*s were then given 10 preliminary training trials at 1 trial per day in a 48-in. alley painted flat gray throughout, to train *S*s to push a door similar to the ones in the arms of the T maze. The last 18 in. of the alley served as a goal box and was separated from the alley by a guillotine door. The *S*s were required to push the door which was mounted in a V-shaped groove similar to those in the T maze. Those *S*s failing to meet this criterion for two successive days were discarded. Reinforcement was 10 sec. of drinking.

After preliminary training, *S*s were randomly assigned to groups. One group of 36 *S*s was run in the choice point maze and given a 67:33 reinforcement schedule (Group C67), while a second group ($n = 36$) received an 80:20 reinforcement schedule (Group C80). The *S*s in Group I and L ($n = 15$ each) were given reinforcement on a 67:33 schedule. Reinforcement schedules were assigned as in Exp. I. Reinforcement was 10 sec. of drinking. Each group was further divided into squads and *S*s within squads were run in a different random order each day. The *S*s were given 1 trial per day. The first four trials were reinforced forced trials as in Exp. I.

The same procedure used in Exp. I was used, except here a response was defined as being the first door in the arm of the maze which was pushed approximately $\frac{1}{4}$ in. When *S* responded to the nonreinforced side, both doors were left in place until *S* corrected and responded to the other door. After completing 10 sec. of drinking, *S* was returned to its home cage. The deprivation schedule and the discard criterion of Exp. I were used here.

The *S*s were run to "asymptote" followed by an additional 20 trials. As a result, the choice point *S*s were given 111 trials, while the I and L Group *S*s were given 122 trials.

In the last month of the study, an epidemic of unknown origin swept the colony and resulted in the loss of 24 *S*s. In addition, 3 *S*s were discarded for failing to run on the early trials, and 2 died during the early part of the study, leaving 30, 25, 8, and 10 *S*s in Groups C67, C80, I, and L, respectively.

The side where the larger proportion of reinforcement was available for each *S* was defined as the A_1 side.

The same method used in Exp. I was used here to estimate the parameters of the various models. Because of the small *n* in Groups I and L, PM predictions were not made for these groups, and only the data from Groups C67 and C80 have been emphasized. The mean A_1 proportion over Trials 92–111 for Groups C67 and C80 and over Trials 102–122 for Groups I and L was taken as an estimate of the asymptotic response probability.

Results

The mean proportion of A_1s during the last 20 trials was .71, .85, .68, and .79 for Groups C67, C80, I, and L, respectively. Statistical analyses failed to reveal any significant differences between each group and its respective reinforcement proportion, $t < 1.63$, $p > .05$. The distributions of the proportions of A_1s during the asymptotic block are presented in Table 3 and reveal that more than half of the *S*s were absorbed on the A_1 side, but only three, across all groups, on the A_2 side.

In Table 4 the parameter estimates for the various models are presented, and in Table 3 are the distributions of the proportion of A_1s during the last 20 trials for the stat rats. Kolmogorov-Smirnov tests were performed comparing the cumulative frequency distributions of real rats to stat rats and revealed significant differences for Groups C67

TABLE 3

Asymptotic Mean Proportions of A_1s ($\bar{P}_\infty(A_1)$) and Distribution of the Proportion of A_1s during the Last 20 Trials in Exp. II

Item	Frequencies of A_1 proportions				
	$\bar{P}_\infty(A_1)$	.8–1.0	.5–.79	.21–.49	0–.2
Group C67					
Real	.71	13	12	3	2
Equal θ	.64	6	23	1	0
2θ	.85	22	8	0	0
ID	.91	24	6	0	0
BW	.70	13	12	3	2
PM	.59	8	18	3	1
Group I					
Real	.68	4	1	3	0
Equal θ	.61	1	4	3	0
2θ	.92	8	0	0	0
ID	.97	8	0	0	0
BW	.65	3	2	3	0
Group L					
Real	.79	8	1	0	1
Equal θ	.66	2	8	0	0
2θ	.85	10	0	0	0
ID	.88	9	1	0	0
BW	.69	5	2	0	3
	$\bar{P}_\infty(A_1)$	.9–1.0	.5–.89	.21–.49	0–.2
Group C80					
Real	.85	17	7	1	0
Equal θ	.81	6	19	0	0
2θ	.95	21	4	0	0
ID	.91	22	3	0	0
BW	.85	15	9	1	0
PM	.74	9	12	3	1

TABLE 4

Parameter Estimates for Exp. II

Item	Group			
	C67	C80	I	L
Equal θ				
θ	.060	.065	.086	.033
2θ				
θ_1	.075	.051	.080	.073
θ_2	.010	.005	.009	.002
ID				
θ	.050	.043	.078	.066
BW				
θ_1	.078	.067	.068	.082
θ_2	.052	.040	.043	.042
PM				
C	.12	.20	—	—
N	3	2	—	—

and C80 for the unequal θ and ID models and for Group C80 for the equal θ model, $K_D \geq 11$, $p < .05$. The comparison for Group C67 for the equal θ model, $K_D = 9$, $p > .05$, and for Group C80 for the PM, $K_D = 10$, $p > .05$, just missed significance. The comparisons for Group C67 for the PM, $K_D = 5$, $p > .05$, and for both groups for the BW model, $K_D < 3$, $p \geq .05$, failed to yield significance.

As in Exp. I, 44 statistics were used to explore the fine detail of the data. These analyses (see Robbins, 1969 for details) revealed that 6 and 10 of the 44 comparisons for Groups C67 and C80 for the BW stat rate and 1 and 8 for the PM rats were statistically significant. In general, BW stat rats tended to repeat responses less so than real rats, while most of the PM differences were related to the larger number of real rats absorbed.

Discussion

Theoretical analyses of Exp. II revealed that the asymptotic frequency distributions were adequately described by the BW model, and less so by the PM. The PM, however, did describe the fine detail for Group C67 well, but shortcomings were noted for Group C80. In contrast, the BW model fell short of an adequate description of the fine detail for both groups. Thus Exp. II failed to discriminate between these two models. The inadequacies of the BW model in Exp. II may be because correction occurred in the presence of choice point stimuli, while in Exp. I it did not. A secondary reinforcement mechanism may also account for the discrepancies if one assumes that primary reinforcement affected the corrected response only and secondary reinforcement the initial choice response in Exp. I but not in Exp. II.

The results of both studies suggest that absorption rather than matching may be the typical outcome for probability learning in the correction T maze. In addition, the data of Exp. I suggest that *S*s may develop stereotyped behavioral modes which may be extremely difficult to disrupt. If the BW model is correct, the difficulty in disrupting choice behavior is due to increases in the probability of the initial choice response. This suggests that "fixation" of choice behavior may be treated as a special case of the BW model.

REFERENCES

BITTERMAN, M. E. Phyletic differences in learning. *American Psychologist,* 1965, **20**, 396–410.

BITTERMAN, M. E., WODINSKY, J., & CANDLAND, D. K. Some comparative psychology. *American Journal of Psychology,* 1958, **71**, 94–110.

BRUNSWIK, E. Probability as a determiner of rat behavior. *Journal of Experimental Psychology,* 1939, **25**, 175–197.

BUSH, R. R., & MOSTELLER, F. Stochastic models for learning. New York: Wiley, 1955.

BUSH, R. R., & WILSON, T. R. Two-choice behavior of paradise fish. *Journal of Experimental Psychology,* 1956, **51**, 315–322.

CARTERETTE, T. S. Terminal probability of a position response as related to three correction procedures. Unpublished master's thesis, Indiana University, 1957.

ESTES, W. K. Toward a statistical theory of learning. *Psychological Review,* 1950, **57**, 94–107.

ESTES, W. K. Of models and men. *American Psychologist,* 1957, **12**, 609–617. (a)

ESTES, W. K. Theory of learning with constant, variable or contingent probabilities of reinforcement. *Psychometrika,* 1957, **22**, 113–132. (b)

ESTES, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford University Press, 1959.

PARDUCCI, A., & POLT, J. Correction vs. noncorrection with changing reinforcement schedules. *Journal of Comparative and Physiological Psychology,* 1958, **51**, 492–495.

ROBBINS, J. Probability learning in the correction T maze. Unpublished doctoral dissertation, City University of New York, 1969.

WEINSTOCK, S., NORTH, A. J., BRODY, A. L., & LOGUIDICE, J. Probability learning in the T maze with noncorrection. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 76–81.

WRIGHT, R. L. D. Motivational effects in probability learning. *Psychonomic Science,* 1967, **7**, 329–330.

(Received February 3, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 121–128

GROWTH OF A PERCEPT AS A FUNCTION OF INTERSTIMULUS INTERVAL

GERALD M. MURCH [1]

Portland State University

Words were presented tachistoscopically 1, 3, 5, and 10 times, with interstimulus intervals (ISIs) of 50, 100, 250, and 1,000 msec. (Exp. I) and 10, 25, 50, and 75 msec. (Exp. II). The increase in probability of correct identification as a function of frequency displayed the same slope for all ISIs, but overall height of the curves varied with the interval used. As the ISI decreased from 1,000 to 50 msec., a decrement in performance was observed and explained as the result of a reduction in scanning time. This was empirically supported in Exp. III in which a fixation point was added. An improvement at short intervals was explained via a hypothesis of visual summation. Moving the word relative to the fixation point (Exp. IV) failed to reduce the improvement in performance at the shorter ISIs. It was concluded that visual summation is not a function of the primary receptors.

Recently, Haber and Hershenson (1965) addressed the problem of the effects of the repeated exposure of a meaningful word on the probability that the word would be correctly identified. They noted that studies of word recognition thresholds, using the method of limits, apply a number of independent variables, e.g., set, word frequency in print, word length, etc., while confounding the variables of exposure duration and exposure frequency. Using several durations, Haber and Hershenson varied the frequency of the exposure of common seven-letter words and showed the relationship between identification probability and frequency of exposure to be a negatively accelerating function approaching an asymptote after approximately 10 presentations of the word.

Recognition accuracy in the Haber and Hershenson (1965) study could be predicted with several empirical functions. Of these functions, the best fit of the data points for the varying durations of exposure (threshold, threshold plus 5 msec., threshold plus 10 msec.) was $p_n = A - B/n$, where A is the maximum probability that the word would be perceived, B is the probability of incorrect identification on the occasion of the first exposure, and n is the number of exposures.

While the work of Haber and Hershenson (1965) provides a useful description of the relationship between recognition probability and presentation frequency, it fails to incorporate the variable of the interval between successive stimulus presentations (interstimulus interval or ISI). The importance of this variable in studies observing an interaction between temporally separate perceptual inputs is emphasized by recent work on *critical durations* in visual acuity tasks. Kahneman (1964, 1966) has traced the length of time over which the time–intensity reciprocity (Block's law) is shown to hold. The research on visual masking (Eriksen & Collins, 1965) and on short-term perceptual memory (Sperling, 1963) has indicated that critical temporal periods exist in the visual mechanism within which input and energy appear to summate or interact.

EXPERIMENT I

The purpose of Exp. I was to observe the relationship between recognition accuracy and frequency of exposure as a function of the interval between successive presentations of the stimulus words. In the study by Haber and Hershenson (1965), the ISI was held constant at approximately 10 sec.[2]

---

[1] The author wishes to express his appreciation to Dianna Anderson for her aid in running *S*s and evaluating the experimental results and to Robert E. Jones, Jr., and Barry F. Anderson for comments on the manuscript.

Requests for reprints should be sent to Gerald M. Murch, Department of Psychology, Portland State University, Portland, Oregon 97207.

[2] R. N. Haber, personal communication, October 15, 1965.

A further difference between Exp. I and the earlier work of Haber and Hershenson (1965) relates to the visual angle of the stimulus array. Their stimuli (English words) subtended small visual angles (4°). In the present study, a *scanning* task was introduced in that the words subtended visual angles of 13°. Whereas in the study by Haber and Hershenson all elements of the words could be perceived simultaneously, in Exp. I a scanning or refixation would be required in order to sample all the elements.

Method

The 220 most frequent seven-letter words from the Thorndike-Lorge (1944) list were selected. The list excluded proper names. All words were printed on white 4 × 6 in. file cards with a ½-in. LeRoy stencil.

The words were presented in a two-field mirror tachistoscope (Gerbrands Model T-2B-1). Field 1 was always illuminated except for the duration of the presentation of the word in Field 2. Field 1 contained two penciled horizontal lines (1 in. apart). The luminance of the fields was held constant at 4.6 ftc.

Four exposure frequencies (1, 3, 5, and 10) and four interstimulus intervals (50, 100, 250, and 1,000 msec.) were used. Each *S* was shown 10 words at each of the 16 Frequency × ISI conditions, the order of conditions being randomized across *S*s.

The *S*s ($N = 16$) were unpaid volunteers from the author's introductory psychology class. At the start of each experimental session, the deck of 220 words was mixed and 10 words were selected for the establishment of the presentation duration for *S*. The *S* was informed that seven-letter English words would be shown several times. Each time the signal light, mounted below the viewer of the tachistoscope, came on, *S* was to try to identify the word presented. Each word was presented 10 times, with an ISI of 1,000 msec. The duration was set at 1 msec. on Trial 1 and increased by 1 msec. after each 10 presentations until *S* correctly identified the word. A duration for each *S* was desired that would allow the probability of word identification after 10 presentations with an ISI of 1,000 msec. to approach 1.0. Therefore, the lowest duration at which at least 6 of the 10 words were correctly identified was taken as the presentation duration for Exp. I for the particular *S*.

For Exp. I, *S* was given an answer sheet containing the numbers 1–160. After each number were seven blank spaces in which *S* was to record his answers. He was told to record what he saw, even if it was not a complete word. After each word, *E* gave *S* a code number to be entered in a special column to the left of the number of the trial. The presentation of each word for the specific Frequency × ISI condition occurred when *S* depressed a Morse key located below the tachistoscope. That is, the stimulus appeared either 1, 3, 5, or 10 times at one of the four ISIs. Naturally, the condition of one word presentation at each ISI represented a control condition. The signal lamp below the viewer came on at the end of the presentation, at which time *S* was to record his answer. While *S* recorded his perception of the word presented, *E* changed the word, selecting the new word from the mixed deck of words. After 10 words, a new Frequency × ISI condition was set.

Results

An analysis of variance of the number of words correctly identified for the three frequency conditions (3, 5, and 10 presentations of each word) and the four ISIs revealed a significant effect for frequency, $F\ (2, 45) = 11.26$, $p < .01$, and for ISI, $F\ (3, 129) = 12.04$, $p < .01$.

The four control conditions of one presentation of the word at each of the ISIs were compared with a Friedman test. The resulting value of $\chi^2_r = 3.51$ was not significant at the .10 level.

Figure 1 presents the relationship between presentation frequency and the probability of correct word identification over all ISI conditions. As the analysis of variance indicated no ISI × Frequency interaction, the data for the ISIs of 100, 250, and 1,000 msec. have been summed for this curve. Also included in Fig. 1 is the curve predicted by the function offered by Haber and Hershenson (1965). That function overestimates the slope of the observed function obtained in Exp. I for the summed intervals.

The data for the 50-msec. ISI were not included in the summed curve as it showed a marked deviation from the other three ISIs. The mean number of words correctly identified for each frequency was compared for each ISI condition by Duncan's multiple-range test. The means for the 50-msec. ISI deviated significantly ($p = .01$) from all other ISI means for frequencies of 5 and 10 presentations.

The effects of ISI are shown in Fig. 2. As can be observed, reductions in ISI appear

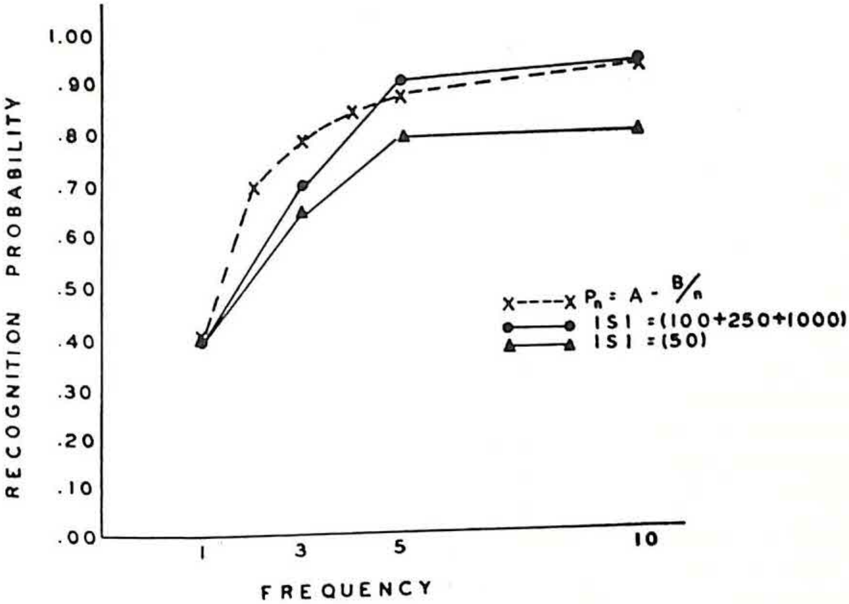


FIG. 1. The probability of correct word identification as a function of frequency of presentation for ISIs of 100, 250, and 1,000 msec. combined, an ISI of 50 msec., and the predicted relationship between recognition probability and presentation frequency given by Haber and Hershenson (1965).

to result in a monotonic decrease in recognition accuracy over the intervals used.

Discussion

Under the conditions of Exp. I, the relationship between identification probability and frequency of exposure is found to be a negatively accelerated function. This result is comparable to the data obtained by Haber and Hershenson (1965). The slight overestimation by their function can probably be attributed to the large visual angle subtended by the stimuli in Exp. I. The observation of a reduction in recognition accuracy with decreasing ISI would be expected in a task such as that imposed on *S* in Exp. I. A 5-in., seven-letter word presented at a distance of 21 in. subtends a horizontal visual angle of 13.6°. In order to accurately perceive the word, *S* must fixate the first part of the word and then refixate the last part of the word. Assuming that such refixations can occur as rapidly as once each 300 msec., additional information about the word could only be obtained with total presentation times (Frequency × ISI) greater than 300 msec. As the interval decreases, *S* has less time to scan the array and recognition accuracy decreases.

The data from Exp. I can be adequately explained with the assumption of information loss due to a reduction in scanning time at shorter ISIs. The purpose of Exp. II was to observe the effects of shorter ISIs on recognition accuracy as a function of the frequency of presentation.

EXPERIMENT II

A number of experiments have demonstrated the existence of a perceptual time unit somewhere in the region of 50–100 msec. (White, 1963). Although a variety of interpretations have been advanced, the empirical evidence speaks for a process of visual summation in which successive visual stimuli falling within the critical time period interact. Such critical durations relate on one hand to the time–intensity reciprocity, in which visual responses are found to be dependent on the total energy of the stimulus ($E = I \times t$, where I equals illumination and t the duration of the stimulus). Kahneman (1964) demonstrated a nonmonotonic relationship between stimulating energy and acuity with regard to the critical duration within which the reciprocity law is shown to hold. As luminance increased, the critical duration decreased from 200 msec. to somewhere between 50 and 100 msec. Further increases in energy resulted in an increase in the critical duration to about 300 msec. In a recent study by Eriksen and Collins (1967), two patterns of dots, which when presented simultaneously produced a legible nonsense syllable, were presented at varying interstimulus intervals ranging from 0 to 500 msec. The probability of the correct identification of the syllable was observed as a function of the ISI. At very short ISIs (below 25 msec.), performance approached 100% correct identification. As the ISI increased, performance decreased, reaching an asymptote somewhere in the region of 100–300 msec.

Thus on one hand, some tasks (Kahneman & Norman, 1964) result in complete reciprocity for illumination and duration and no reduction in recognition performance over critical durations up to 300 msec. (Kahneman, 1964). Other tasks, such as that employed by Eriksen and Collins (1967), yield a monotonic decrease in recognition accuracy in which performance is inversely related to ISI and reaches an asymptote between 100 and 300 msec.

Method

The method was the same as Exp. I with the exception of the ISIs used (10, 25, 50, and 75 msec.). All other aspects of the study, including the measurement of threshold, paralleled Exp. I. The *S*s ($N = 16$) were again unpaid volunteers from an introductory class.

Results

An analysis of variance for the number of words correct for the three frequencies of presentation (3, 5, and 10) revealed a nonsignificant value, $F\ (2, 45) = 1.10$, $p > .10$. The ISI variable was significant, $F\ (3, 129) = 6.46$, $p < .01$, and the Frequency × ISI interaction was not significant, $F\ (12, 129) = 1.51$, $p < .10$. The insignificant value for the frequency condition was found to be due to the deletion of the control condition of one presentation at each ISI. In a subsequent 4 × 4 analysis, the value for frequency was found highly significant, $F\ (3, 45) = 17.28$, $p < .01$.

Again the control conditions of one word presentation at each ISI were compared with a Friedman test. The resulting value of $\chi^2_r = 4.03$ again failed to reach a .10 level of significance.

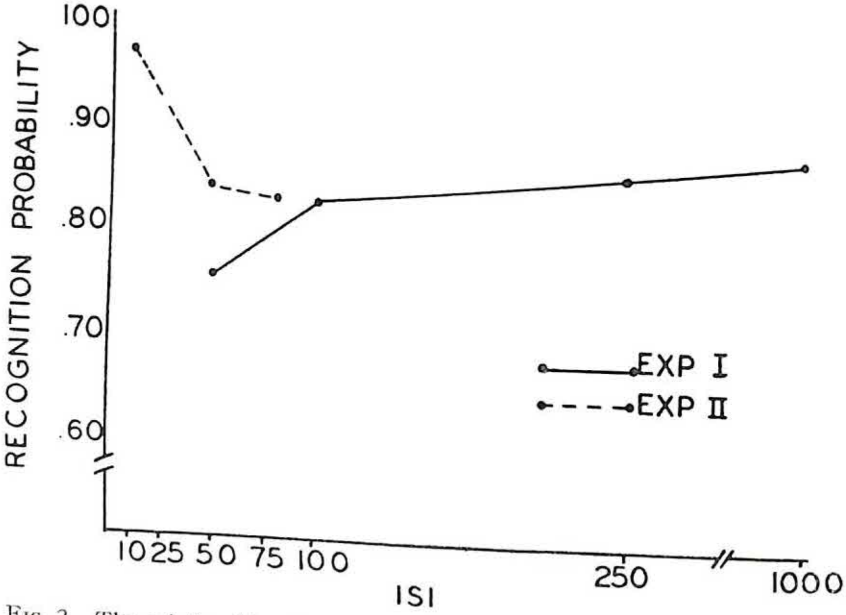


Fig. 2. The relationship of ISI to recognition probability for all frequencies combined in Exp. I and II.

Figure 2 presents the relationship between recognition probability and ISI for Exp. II. Also included in Fig. 2 are the data from Exp. I.

Discussion

The probability of correct identification of the words was shown to decrease from 1,000 to 50 msec. (see Exp. I). For the very short ISIs in Exp. II (10 and 25 msec.), an increase in the number of correct identifications resulted. This observation confirms the observations of Eriksen and Collins (1965) in the masking situation and Eriksen and Collins (1967) in the successive presentation of dot patterns forming a nonsense syllable. The summation effect decreases monotonically from zero-delay periods to an asymptote around 100 msec. It does not fit with the expectations of the critical durations for the time–intensity reciprocity function, which would predict no reduction in response accuracy within the confines of the critical duration (Kahneman & Norman, 1964). Such a discrepancy between functions for form perception and for brightness summation with regard to a critical duration has been observed by Schurman, Eriksen, and Rohrbaugh (1968). They concluded that perceived brightness within a critical duration is solely a function of the amount of energy present within the critical duration. For form perception, critical durations were found only when the light energy was continuously distributed over the interval of stimulation.

The data of Exp. II, then, appear compatible with a visual summation hypothesis in which the probability of correct word identification decreases linearly as the interval between successive stimulus presentation is lengthened.

The large discrepancy between recognition probabilities in Exp. I and II for the 50-msec. ISI seem worthy of note. If one considers *S*'s task in Exp. I as compared to Exp. II, it would appear reasonable to assume that *S* would adopt a different strategy in each study to maximize the probability of identifying each word. In Exp. I, all but one of the ISIs were longer than 100 msec. An efficient strategy for the long ISIs would involve fixating both the front and back parts of the word. On those trials using the 50-msec. ISIs, this strategy would be most inefficient as *S* would have time to fixate only one portion of the word. In Exp. II, however, all ISIs were below 100 msec. In this case, *S* would never have time to fixate on various sections of the word and would use a single fixation point in the center of the word as the most efficient strategy.

EXPERIMENT III

The explanation of the results of Exp. I and II as a dual process dependent on a reduction in scanning time at longer intervals and visual summation at shorter intervals can be empirically examined. This was the task of Exp. III and IV. By including a fixation point and instructing *S* to fix his attention on that point during the entire presentation, scanning could be reduced. Also, the inclusion of a fixation point, due to the increased probability of successive stimulus presentations falling on congruent retinal areas, might facilitate summation.

Method

Exp. III deviated from Exp. I in only two aspects. First, instead of two penciled horizontal lines in Field 1 of the tachistoscope, a fixation point ($\frac{1}{8}$-in. X) was centered in the field. The word was centered on the point. Second, a combination of ISIs from Exp. I and II was used (25, 50, 100, and 500 msec.). All other aspects of the experiment paralleled Exp. I.

Results

The results of Exp. III are presented in Fig. 3. The figure relates recognition probability to ISI over all frequency conditions. The combined data from Exp. I and II are provided to give a base line against which the results of Exp. III can be judged.

The 3×4 analysis of variance revealed an $F\ (2, 45) = 4.43$, $p < .025$, for frequency and an $F\ (3, 129) = 12.59$, $p < .005$, for the interval between presentations. Again the Frequency × Interval interaction failed to attain significance, $F\ (12, 129) = .73$, $p > .05$.

In order to test the effects of a fixation point, a within-groups analysis of variance (Lindquist, 1953) across Exp. I and III was conducted. The results showed a significant interaction between Exp. I and III and ISI, $F\ (2, 60) = 10.91$, $p < .001$.

A Friedman test conducted for the control condition of one presentation of each word for each ISI yielded a value of 3.32, which does not reach a .10 level of significance.

Discussion

The inclusion of a fixation point resulted in lower recognition probabilities at the longer ISIs (Exp. III) as compared to a condition in which *S* was not required to maintain a strict fixation (see Exp. I and II). This result is compatible with the assumed scanning process used to explain the improvement at longer ISIs in Exp. I and II. By requiring *S* to maintain a strict fixation, scanning behavior was reduced such that longer intervals had no facilitatory effect.

EXPERIMENT IV

Experiment IV tested the assumed process of visual summation used in the explanation of the improvement in recognition accuracy at short ISIs (10, 25, and 50 msec.). It was hypothesized that changing the position of the word relative to the fixation point during each ISI should reduce the opportunity for visual summation at the level of primary receptors of the retina. This hypothesis follows from the data of Eriksen (1966), who demonstrated independence of successive presentations of the same or different forms projected on different portions of the fovea, even when the interval between presentations was less than 1 msec.

Method

The method of Exp. IV paralleled Exp. III in that ISIs of 25, 50, 100, and 500 msec., frequencies of presentation of 1, 3, 5, and 10, and a fixation point were used. Other aspects, such as the establishment of the presentation duration, words, and design, were the same as in the previous studies. In Exp. IV, however, the card holder for the word was mounted on the presser foot of a Singer sewing machine placed in a light-shielded cabinet built behind Field 2 of the tachistoscope. Independent judges were unable to detect any difference in the appearance of the words when flashed while mounted in front of the sewing machine as compared with the presentation in the card holder of Field 2 in the prior experiments. The sewing machine was activated .5 sec. before the first presentation of the word and continued to run during the ISI between presentations. Naturally, the machine was inactive with a presentation frequency of 1. The word was moved 2 in. in a vertical direction at a constant speed of 12 cps. With this speed, the position of the word varied between ¼ and 1 in. relative to its position on the preceding flash. The mean variation was ½ in. The duration of each word established for each *S* ($M = 2.3$ msec.) was short enough to "freeze" the word at the position in the continuous cycle of movement set up by the sewing machine.

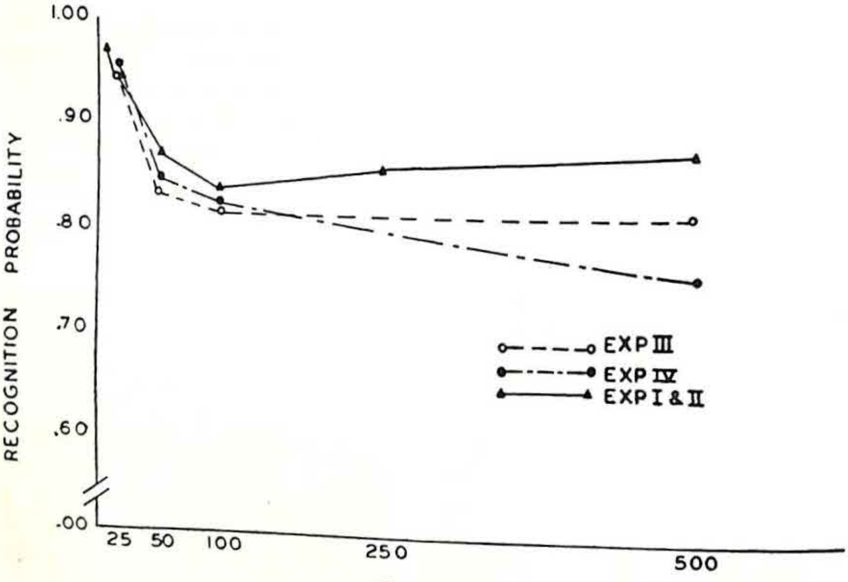


FIG. 3. The relationship of ISI to recognition probability for all frequencies combined in Exp. I and II, III, and IV.

Results

The effect of ISI, independent of frequency, for Exp. IV is illustrated in Fig. 3. The combined data from Exp. I and II as well as the results of Exp. III are included to provide a comparison of the effect of moving the position of the word during the ISI.

The 3×4 analysis of variance resulted in a significant value for frequency of word presentation, F $(2, 45) = 7.64$, $p < .005$, as well as for the effect of ISI, F $(3, 129) = 10.02$, $p < .005$. The Frequency × ISI interaction was not significant, F $(12, 129) = .89$, $p > .05$.

The effect of moving the word relative to the fixation point can be assessed by a comparison of Exp. III and IV. The results showed neither a significant interaction between Exp. III and IV and frequency, F $(2, 60) = .65$, $p > .05$, nor between Exp. III and IV and ISI, F $(3, 90) = .43$, $p > .05$. Comparing all three experiments (I, III, and IV) revealed the expected significant Experiments × ISI interaction, F $(4, 90) = 7.36$, $p < .001$, and a failure of the Experiments × Frequency interaction to attain significance, F $(4, 90) = 1.51$, $p > .05$.

The control conditions of one word presentation at each frequency were compared via a Friedman test. Again the four conditions did not differ at the .10 level.

Discussion

The results of Exp. IV indicate an improvement in recognition accuracy even though the stimulus words were projected on different parts of the fovea on successive presentations. An interpretation of these data in line with a visual summation hypothesis emphasizing primary receptors of the visual system does not seem plausible. At first consideration, these results would appear compatible with a process of summation occurring at a higher visual center, such as has been suggested by the physiological work of Hubel and Wiesel (1962, 1965). Kahneman (1966) suggested the possibility of explaining temporal summation of form in terms of a central mechanism and provided data (Kahneman, 1964) to support the assumption that the primary receptors are not responsible for temporal summation. A recent study by Kahneman, Norman, and Kubovy (1967) observed the extent of temporal summation for a form (Landolt C) under conditions of monocular and binocular viewing, the binocular presentation being either simultaneous or successive. Their results showed perfect binocular summation in the binocular-successive condition up to 300 msec. and up to 160 msec. for the binocular-simultaneous and monocular conditions. They conclude:

> This indicates that the structure which is responsible for binocular summation can utilize with complete effectiveness information arriving from the two eyes during this period. Therefore, the deterioration of performance beyond 160 msec. in the monocular and binocular simultaneous conditions cannot be due to limitations of this structure or any subsequent stage of visual processing. The decrement of simultaneous binocular vision for durations exceeding 160 msec. must then be due to the weakening of messages arriving from a more peripheral station along the visual pathway [p. 326].

If it is possible to assume that the underlying process of summation in the Kahneman et al. (1967) experiment was mediated by the same mechanism as the summation effect observed in the present study, the locus of the effect can be neither cortical nor primary. By the process of elimination, the summative effect must occur at the level of the bipolar or ganglion cells, as suggested by Murch (1967) and Kahneman et al. (1967), or in the lateral geniculate body. This latter structure of the visual system has been shown in the cat to display a lesser degree of binocular interaction (Erulkar & Fillenz, 1960) than in the visual cortex (Hubel & Wiesel, 1962).

REFERENCES

ERIKSEN, C. W. Independence of successive inputs and uncorrelated error in visual form perception. *Journal of Experimental Psychology*, 1966, **72**, 26–35.

ERIKSEN, C. W., & COLLINS, J. F. Reinterpretation of one form of backward and forward masking in visual perception. *Journal of Experimental Psychology*, 1965, **70**, 343–351.

ERIKSEN, C. W., & COLLINS, J. F. Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology*, 1967, **74**, 476–484.

ERULKAR, S. D., & FILLENZ, M. Single-unit activity in the lateral geniculate body of the cat. *Journal of Physiology*, 1960, **154**, 206–218.

HABER, R. N., & HERSHENSON, M. Effects of repeated brief exposures on the growth of a precept. *Journal of Experimental Psychology*, 1965, **69**, 40–46.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology,* 1962, **160**, 106–154.

Hubel, D. H., & Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology,* 1965, **28**, 229–289.

Kahneman, D. Temporal summation in an acuity task at different energy levels—a study of the determinants of summation. *Vision Research,* 1964, **4**, 557–566.

Kahneman, D. Time intensity reciprocity in acuity as a function of luminance and figure ground contrast. *Vision Research,* 1966, **6**, 207–215.

Kahneman, D., & Norman, J. The time-intensity relations in visual perception as a function of observer's task. *Journal of Experimental Psychology,* 1964, **68**, 215–220.

Kahneman, D., Norman, J., & Kubovy, M. The critical duration for the resolution of form: Centrally or peripherally determined? *Journal of Experimental Psychology,* 1967, **73**, 323–327.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Murch, G. M. Temporal gradients of responses to subliminal stimuli. *Psychological Record,* 1967, **17**, 483–491.

Schurman, D. L., Eriksen, C. W., & Rohrbaugh, J. Masking phenomena and time-intensity reciprocity for form. *Journal of Experimental Psychology,* 1968, **78**, 310–317.

Sperling, G. A. A model for visual memory tasks. *Human Factors,* 1963, **5**, 19–31.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

White, C. T. Temporal numerosity and the psychological unit of duration. *Psychological Monographs,* 1963, **77** (12, Whole No. 575).

(Received February 3, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 129–135

POSITIVE AND NEGATIVE TRANSFER RESULTING FROM FORMAL SIMILARITY OF STIMULI[1]

WILLARD N. RUNQUIST[2]

University of Alberta

Transfer in A-B, A'-B and A-B, A'-B$_r$ paradigms was related to number and position of identical letters between A and A' terms. Compared with an A-B, C-B condition, positive transfer occurred only when A and A' terms shared two letters. Negative transfer was also restricted to situations where stimuli shared two letters. When letter identity was in the same position for all items, negative transfer did not appear until late in acquisition. The results, in general, do not conform to a stimulus generalization interpretation of transfer and suggest the involvement of possible mediational mechanisms in transfer due to formal similarity of stimuli.

The purpose of the present experiment was to study transfer in the A-B, A'-B and A-B, A'-B$_r$ paradigms as a function of the formal similarity between A and A', the stimulus terms of two consecutively learned paired-associate lists.

Despite extensive empirical investigation of the relationships between transfer performance and similarity, relevant data on this particular problem are singularly lacking. Most studies to date have manipulated meaningful rather than formal similarity (e.g., Bugelski & Cadwallader, 1956; Osgood, 1946; Wimer, 1964), and Underwood (1966) has pointed out that different theoretical processes are likely to be involved in transfer mediated by meaningful relations. Transfer experiments using formal stimulus similarity seem to be restricted to nonverbal stimuli (Dallett, 1962; Gibson, 1941), with one exception. Bruce's (1933) classic study compared an A-B, A'-B paradigm with an A-B, C-B condition. Relative positive transfer was obtained when the trigram stimuli shared the first two letters. Appropriate negative paradigms were not investigated, nor were other letter combinations included. In the only other experiment relevant to this problem, Yum (1931) provided a single "recall" trial after changing the stimuli and found that performance on this test suffered less when fewer letters were different from those used during learning.

Formal similarity among verbal units has traditionally been defined as in the aforementioned studies, i.e., in terms of number and position of identical letters. This definition is certainly reasonable since it has been shown that at least with trigrams, rated similarity is almost perfectly correlated with number of identical letters and is virtually independent of position unless the locus of the letter is changed (Runquist & Joinson, 1968).

In the experiments reported here, selected conditions varying the locus and number of identical letters between stimulus terms in successive lists were employed in both positive (A-B, A'-B) and negative (A-B, A'-B$_r$) transfer conditions. In some conditions the locus of identity was the same for each transfer pair, while in other conditions the locus varied for different pairs. Relative transfer under these conditions has several implications for contemporary transfer theory, but discussion of these issues is postponed until the actual results have been presented.

METHOD

Design and materials.—The *S*'s task was to learn to press one of six buttons arranged in a linear array when one of six verbal stimuli was presented on a visual display. This task has the advantage of

---

[1] This study was supported in part by Grant APB-88 from the National Research Council of Canada and in part by Grant GB-6166 from the National Science Foundation. Lillian Bowes supervised the collection of the data. The *E*s were Edward Davies, Richard Upshall, James Wheatley, and Dexter Harker.

[2] Requests for reprints should be sent to Willard N. Runquist, Department of Psychology, University of Alberta, Edmonton 7, Alberta, Canada.

eliminating transfer of response learning in the two paradigms under investigation, as the "response population" is continuously present. Each *S* performed this task with two sets of six stimuli selected and assigned to response buttons to fit one of nine conditions.

All stimuli were CVC trigrams of less than 25% association value (Archer, 1960). One set of six trigrams served as the second set (List 2) for all *S*s. In all sets, intralist letter repetitions were as low as possible. Each stimulus began and ended with a different consonant, although two consonants were duplicated in the set. In List 2 each of three vowels was used twice, although in other lists vowel duplication necessarily varied somewhat in order to construct various paradigms.

Groups were defined in terms of different first lists. The "control" list fitted an A-B, C-B paradigm, with no letters in common with any items in List 2. In the list designated HE-1 (heterogeneous rules, one letter), each stimulus shared one letter with a corresponding stimulus in List 2. For two pairs it was the initial letter, for two pairs it was the vowel, and for two pairs it was the third letter. Another list, designated HE-2, was constructed so that each item shared two letters with one of the List 2 items. Two pairs shared first and second letters, two pairs shared first and third letters, and two shared second and third letters. The remaining two lists, designated HO-1 and HO-2, were also constructed so that one or two letters, respectively, were shared with items in List 2. In the case of HO-1, however, it was the first letter for all pairs. In the case of HO-2, it was the first two letters in all pairs. The latter two conditions were selected from six possible relations because they had been shown to produce the most intralist interference (Runquist, 1968).

In each case except the control, the two main transfer paradigms were created by assigning items sharing letters to the same button in List 1 and List 2 to produce the A-B, A′-B paradigm and by assigning them to different buttons to make an A-B, A′-B_r paradigm. Thus, there were nine conditions: the control group (A-B, C-B) and the four transfer groups (HE-1, HE-2, HO-1, and HO-2), arranged with both relevant pairing and re-pairing of terms sharing letters. All stimuli are shown in Table 1.

There are two facets of this design which deserve comment. Traditionally, A-B, C-D (new stimulus and response terms) provide the base line from which transfer is measured. This does not since the interest here lies in relative transfer effects seem necessary or appropriate in the present case resulting from increasing similarity of stimulus terms. Thus, while both positive and negative paradigms converge on A-B, C-B, no statement can be made comparing the amount of transfer in the two directions. One may only make conclusions with respect to the relative effects of increasing similarity.

TABLE 1

STIMULUS MATERIALS

Control	First list				Second list
	HE-1	HE-2	HO-1	HO-2	
RUJ	XAT	XET	QAZ	QEZ	QET
LIW	KIH	KYG	KIH	KYG	KYB
XAP	ZUF	ZAS	ZUN	ZOV	ZOS
HIF	RUV	MIV	MIW	MEJ	MEV
WUX	COJ	GOQ	VUF	VOF	VOQ
GAC	JYX	NYZ	NAJ	NYH	NYD

The second matter of concern is the use of only one source of similarity in the HO conditions. Ideally, all three possible letter identity rules should have been used for both single- and double-letter similarities, thus allowing a direct comparison between HE and HO conditions.

However, transfer designs demand identical second lists for all conditions, and in the present circumstances no lists could be constructed which kept this restriction with all of the possible identity relations. No alternative design could be arranged which did not involve a prohibitive number of conditions and lists. Consequently, it was decided to forego this comparison and use these combinations for HO conditions which were most likely to produce transfer.

Apparatus and procedure.—The apparatus has been completely described elsewhere (Runquist, 1966, 1968). The *S*s were seated in booths within which were located sloping response panels containing the six buttons. Immediately above each button on a vertical extension of the panel was a small light. Protruding from the wall above the booth was a 3 × 5 in. screen on which the trigrams were presented. A system of relays controlled by punched paper tape selected stimuli. A cam driven by a synchronous motor controlled presentation rate. Responses were recorded on an event recorder. Generally, two *S*s were run simultaneously in the same condition.

All learning was by the anticipation method. Each item was presented for 3 sec., of which 1.54 sec. was anticipation interval, during which time *S* tried to press the correct button. During the succeeding .98 sec., the light above the correct button was lighted and responses were not recorded.

For the remaining .48 sec., both stimulus display and feedback lamp were off. Although the order of presentation was arranged so that each item occurred once in each block of six presentations, no interval or other indication was given to *S* that one "trial" had ended and that another had begun. For *S* it was simply an unbroken sequence of stimulus presentations. Five different orders of presentation were used.

Twenty repetitions of each stimulus were given on List 1 and 12 trials on List 2. Approximately 1 min. separated the two lists.

Subjects.—The *S*s were 180 introductory psychology students who served as part of a course requirement. As they appeared at the laboratory, they were assigned to conditions according to a scheme which randomized the order of conditions within blocks containing each condition. All *S*s were run by trained undergraduate *E*s.

RESULTS

List 1 learning.—Performance on List 1 does not provide an unconfounded estimate of the equivalence of the various groups since the lists learned also differed for these groups. Nevertheless, it does provide an assessment of terminal acquisition performance on these lists. The mean number of correct responses on the last five trials is shown in Table 2 for all nine groups. Only the HO-2 list produced performance sufficiently discrepant to cause any difficulty in interpreting transfer performance. Both of the groups which learned this list as List 1 were performing more poorly at the end of acquisition. In the case of the relevant response groups, $F\ (4,\ 95) = 1.42$, $p < .25$, and in the re-pairing groups, $F\ (4,\ 95) = 6.26$, $p < .01$. In one sense, however, significance is not an issue. More important is the bias introduced in transfer comparisons. Although covariance analysis could theoretically be used to adjust transfer data, the regression slopes were heterogeneous, making such an analysis impossible. Consequently, it is simply necessary to take the bias into account in interpreting the transfer results.

TABLE 2

MEAN CORRECT ON LAST FIVE TRIALS OF LIST 1

Group	A-B, A′-B	A-B, A′-B$_r$
Control	19.70	—
HE-1	20.05	20.90
HE-2	18.60	20.35
HO-1	20.65	21.15
HO-2	16.45	13.45

Relevant response conditions (A-B, A′-B).—The left section of Fig. 1 indicates performance on List 2 in blocks of three trials for the four groups in which List 2 responses were paired with relevant List 1 items and for the A-B, C-B (control) condition. As is usually the case, differential performance was restricted to the early trials. Moreover, similarity of the stimulus terms only had a noticeable effect in the conditions in which two letters were shared (HE-2 and HO-2), with the largest effect occurring in the HO-2 group despite the bias introduced by poorer List 1 learning or slower learning *S*s in this group.

An analysis of variance provided some support for the picture shown in Fig. 1. Group means were not significant, $F < 1.00$,

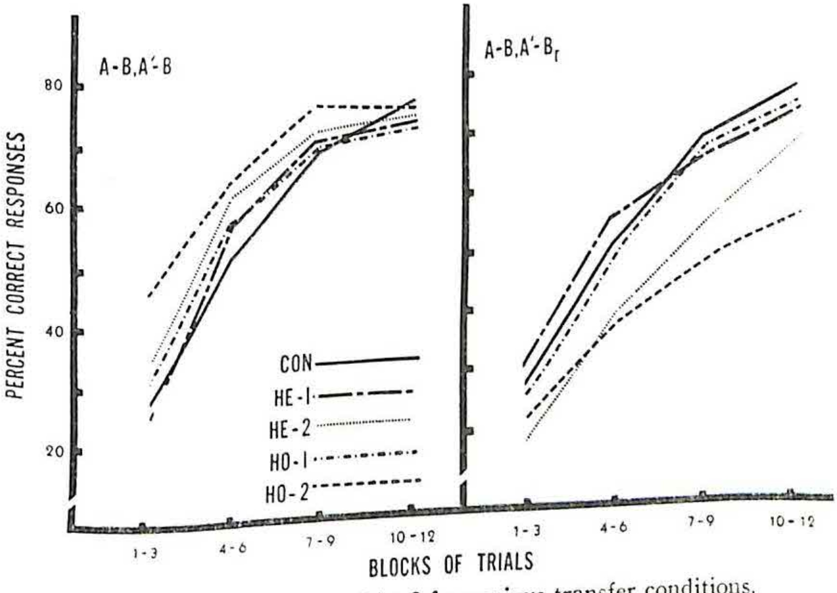


FIG. 1. Performance on List 2 for various transfer conditions.

although the linear component of the Groups × Trials interaction reached acceptable levels, $F\ (4,\ 285) = 3.96$, $p < .01$. Comparisons of each transfer condition with the C-B group showed that only the HO-2 group produced overall facilitation, $t\ (38) = 3.00$, $p < .01$. However, both HE-2 and HO-2 showed significant linear trend differences from that of the C-B group: $F\ (1,\ 285) = 4.65$, $p < .05$, for the HE-2; and $F\ (1,\ 285) = 12.32$, $p < .01$, for the HO-2.

Thus, unless the stimuli shared two letters, there was no direct transfer to List 2 learning resulting from similar letters. Whether the identical letters were in the same position for all pairs or in different positions for all pairs did seem to affect transfer in that the HO-2 group did show a reliable effect, while HE-2 did not.

As previously mentioned, the design does not allow for comparisons between HE and HO conditions, so that assertions concerning the effects of heterogeneity and homogeneity of the similarity relations cannot be made with any confidence.

Nevertheless, indirect evidence may be gleaned from an analysis of items having different position similarities in the HE-2 conditions and from a comparison of performance on these items with comparable items in the HO-2 conditions. If better performance in the HO-2 condition is a result of homogeneity of the transfer relation per se, then items in HE-2 sharing the first two letters should still show less relative transfer. On the other hand, if transfer is equivalent on those items for the two groups, it could be argued that the poorer overall performance in HE-2 is a result of the fact that the other transfer relations included in this condition are less effective. At the very least, this result would not support a conclusion that homogeneous transfer relations produce larger effects.

The analysis was made on the total correct on the first five trials of List 2. Since items were not balanced over button positions, performance on each item was compared with performance on the corresponding item in the C-B group. The difference in total correct responses indicates degree of transfer for those items. For the HE-2 group the overall superiority over C-B was 51 correct responses. The pairs identical in the first two letters accounted for over half of this difference (28 correct responses). Pairs sharing first and third letters produced most of the remaining facilitation (17 correct responses), with pairs sharing vowel and third letter showing little positive transfer (6 correct responses).

The comparable items in the HO-2 condition produced facilitation of 31 correct responses, which corresponds quite closely to the value of 28 in those items in HE-2 group (first two letters identical). The other sets of items in the HO-2 group produced facilitative effects of 38 and 32 correct responses. Consequently, there was no firm evidence for greater transfer when position identity "rules" were the same for all pairs.

Re-paired response conditions (A-B, A'-B_r).—The right section of Fig. 1 shows the acquisition performance of List 2 for the re-paired groups and for the C-B control group. The results parallel the findings with respect to facilitation effects in that increase in interference only occurs when two letters are identical (HE-2 and HO-2). One major difference, however, is that this interference is not restricted to early trials.

Analyses of variance resulted in a significant effect of groups, $F\ (4,\ 95) = 4.15$, $p < .01$, and again the linear component of the Groups × Trials interaction was significant, $F\ (4,\ 285) = 3.76$, $p < .01$. Comparisons of each group with the C-B group showed that mean totals correct for HE-2 and HO-2 were less than that of the C-B, $t\ (38) = 4.95$ and 6.27, respectively. If the slope coefficients of each transfer group are compared with the C-B group, only the HO-2 condition produced any difference, $F\ (1,\ 285) = 10.63$, $p < .01$. This resulted from divergence in acquisition performance over trials.

Although interpretation of the performance of HE-2 is unambiguous, the sampling error or list difference effects on first-list performance of HO-2 makes it impossible to determine whether the generally poorer performance of that group represents a genuine

TABLE 3

MEAN CORRECT FOR BLOCKS OF TRIALS: REPLICATION EXPERIMENT

Group	Correct on List 1	Correct on List 2		
	Trials 16–20	Trials 1–3	Trials 4–6	Trials 7–9
Control	20.5	4.30	11.28	13.78
HO-2	22.3	4.95	10.00	11.69

transfer effect due to similarity. Consequently, the HO-2 condition was replicated along with the C-B condition with 20 *S*s in each group using two new and different sets of lists constructed exactly as in the original experiment. The results are shown in Table 3.

Comparison of the values in the first column provides assurance that this time List 1 performance did not differ between HO-2 and control conditions. Although only nine trials were given on List 2 in this experiment, the same trends are evidenced as before. Initially, HO-2 shows no interference, but by the last block of three trials, performance is considerably lower. Analyses of variance on these data failed to produce a significant overall difference in performance, F (1, 38) = 1.61, but the Groups × Trials interaction was significant, F (2, 76) = 4.22, $p < .05$.

Two other analyses were performed on some of the transfer data from the original experiment. One analysis on the learning of individual pairs in the HE-2 group was made exactly as in the relevant pairing conditions. In the HE-2 group, differences favoring the C-B group held for all three transfer relations. As in the relevant response condition, identical first two letters produced the most transfer (27 incorrect responses). Identical second two letters was next (23 incorrect responses), and the least interference resulted from identical first and third letters (13 incorrect responses).

Further evidence of differential interference was sought in the overt error data, but List 1 intrusions (responses which were correct to the similar stimulus in List 1) did not comprise a greater proportion of overt errors in the HE-2 group than did corresponding response substitutions in the C-B group.

TABLE 4

CONFUSION ERRORS FROM FIRST-LIST STIMULI IN RELEVANT RESPONSE GROUPS

Items	Source	Experimental	Control
HE-1			
2	Vowels	.27	.16
1	Consonants	.24	.51
HE-2			
3	Vowels	.28	.22
3	Consonants	.15	.14
HO-1			
4	Consonants	.10	.06
HO-2			
6	Vowels	.25	.20

Confounding effects.—Inspection of Table 1 reveals that for some lists, transfer relations were confounded by other sources of letter identity within the list. Two sorts of confounding exist: consonants in a different position in an "irrelevant" stimulus and vowel duplication. Not all lists are equally confounded, but in each case the effect should be to decrease positive and increase negative transfer. It is thus possible that failure to find positive transfer in some conditions resulted from such confounding, although failure to find negative transfer in the corresponding re-paired conditions where it would be presumably enhanced argues against such a conclusion.

Nevertheless, an attempt was made to assess the effects in each relevant pairing list by inspecting the source of overt errors. The results are shown in Table 4, which lists the source of negative effects in each group and the proportion of intralist confusion errors among those items sharing letters, along with the proportion of the same error in the control condition where no letter sharing occurred. Clearly, duplication of consonants in different positions had little effect. On the other hand, duplicating vowels does result in small but consistent increases in intrusions.

Whether this interference is responsible for failure of some of the conditions to produce positive transfer is debatable. In the HE-1 group, half of the pairs did not share any vowels with other items, and yet no transfer was observed in those pairs. Furthermore, HO-1 had only consonants confounded. Not only was no relative transfer exhibited in this group, but there was no evidence that items sharing consonants and those not sharing consonants differed. Tentatively, it appears as if basic transfer relations were not affected by these factors.

DISCUSSION

The following major findings emerged from this experiment: (*a*) When response terms were identical, stimulus similarity produced facilitation only when two letters were identical. It was not possible to ascertain whether facilitation was greater when all items shared letters in identical positions as opposed to items sharing different letter positions. (*b*) When response terms were re-paired, similarity produced interference only when stimuli shared two letters. These effects tended to persist during learning, particularly when all items shared letters in identical positions.

One should, of course, be cautious about interpreting these findings in terms of positive or negative transfer. The A-B, C-B condition has been shown to be a negative transfer paradigm when compared with the usual A-B, C-D reference group, although under multiple-choice conditions the effect is quite small (Postman & Stark, 1969). In any event, it is likely that in terms of transfer relative to the C-D type control, A′-B effects are overestimated and A′-B_r effects are underestimated.

Perhaps the most interesting single finding was the negative result obtained with the single-letter conditions. The generality of this result may be debatable, for it may be quite possible to obtain transfer with single letters under appropriate conditions. Nevertheless, the conditions of the present study should have been adequate to produce both positive and negative effects if transfer due to formal similarity is simply a result of stimulus generalization as Underwood (1966, p. 534) suggests. A steep generalization gradient could deal with the "positive" paradigms, but the failure of negative transfer to appear on initial trials of the re-paired paradigms implicates some process other than generalization. Furthermore, there is nothing in the generalization theory to deal with the fact that different sources of identity seem to produce differential transfer effects within both HE-2 conditions.

It may be possible to make some specific assumptions about the generalization process which can deal with these anomalies. On the other hand, it may be a more viable strategy to consider the involvement of other processes. The present experiment does not provide any definitive data suggesting what these processes might be, although some likely possibilities are contraindicated.

Stimulus selection coding (Underwood, 1963) would seem to be a likely candidate, as it has been shown to operate with at least some stimuli of the class used in these studies (Postman & Greenbloom, 1967). If *S* used the identical letters as functional stimuli in List 1, positive transfer could be obtained through direct transfer. This seems unlikely, however, since most findings regarding the selection process indicate fixation on the first letter only, and this condition (HO-1) failed to produce evidence of facilitation or interference. Similar problems occur in the re-paired paradigm in that any reasonable selection strategy on the part of *S* would be likely to result in interference at the outset, which rapidly diminishes.

On this basis, stimulus selection mechanisms do not seem important in this experiment. Of course, the experimental conditions may not have been such as to result in activation of this strategy, so that one should not interpret this to mean that selection is not ever involved in formal similarity-produced transfer. Indeed, it may be reasonable to expect that transfer with one letter identical may only occur when conditions are likely to produce stimulus selection. The "motor" paired-associate task and rapid presentation rate with no intertrial interval used in the present experiment have failed to produce results consistent with a stimulus selection notion in other experiments (Runquist, 1968) as well, so this is a possible solution.

It could also be argued that at least the "positive" effects are due to the elimination of the interference from backward associations presumably involved in the C-B condition. This would mean that the B-A, B-C interference paradigm becomes a B-A, B-A′ paradigm when stimulus terms are similar, and this has long been considered a positive transfer situation. Involvement of this process, at least in these results, seems unlikely. First, no convincing mechanism can be outlined as to how interference or facilitation can take place when *C* or A′ is the presented stimulus term during List 2 learning. Activation of the association presumed to produce interference or facilitation would seem to demand that the correct response be evoked before the relevant backward associate could occur. When it is within the realm of possibility that a backward associate could in some way interact with List 2 performance, it would certainly involve a complex series of events and demand rather strict temporal relations. Secondly, the B-A, B-A′ paradigm usually demands somewhat longer "anticipation" times than those used in the present study if positive effects are to be obtained (Runquist & Marshall, 1963).

One interpretation consistent with the present results is that *S*s use the identical elements or similar sounds to reintegrate or recall the List 1 stimulus and hence its response. Of course, there is no direct evidence of this in the present data, but in view of the fact that other explanations do not consistently conform to the pattern of results, such a process is clearly possible. If so, it would mean that mediating responses are as necessary to account for transfer when similarity is based on formal characteristics of the stimuli as when it is based on meaning.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, 74(10, Whole No. 497).

BRUCE, R. W. Conditions of transfer of training. *Journal of Experimental Psychology*, 1933, **16**, 343–361.

BUGELSKI, B. R., & CADWALLADER, T. C. A reappraisal of the transfer and retroaction surface. *Journal of Experimental Psychology*, 1956, **52**, 360–366.

DALLETT, K. M. The transfer surface re-examined. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 91–94.

GIBSON, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *Journal of Experimental Psychology*, 1941, **28**, 93–115.

OSGOOD, C. E. Meaningful similarity and interference in learning. *Journal of Experimental Psychology*, 1946, **36**, 277–301.

POSTMAN, L., & GREENBLOOM, R. Conditions of cue selection in the acquisition of paired-associate lists. *Journal of Experimental Psychology*, 1967, **73**, 91–100.

POSTMAN, L., & STARK, K. Role of response availability in learning and transfer. *Journal of Experimental Psychology*, 1969, **79**, 168–177.

RUNQUIST, W. N. Intralist interference as a function of list length and interstimulus similarity. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 7–13.

RUNQUIST, W. N. Formal intralist stimulus similarity in paired-associate learning. *Journal of Experimental Psychology*, 1968, **78**, 634–641.

RUNQUIST, W. N., & JOINSON, P. A. Rated similarity of low association value CVC trigrams. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 317–320.

RUNQUIST, W. N., & MARSHALL, M. A. Transfer, synonymity, and anticipatory interval in paired associate verbal learning. *American Journal of Psychology*, 1963, **76**, 281–286.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963.

UNDERWOOD, B. J. *Experimental psychology*. (2nd ed.) New York: Appleton-Century-Crofts, 1966.

WIMER, R. Osgood's transfer surface: Extension and test. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 274–279.

YUM, K. S. An experimental test of the law of assimilation. *Journal of Experimental Psychology*, 1931, **14**, 68–82.

(Received February 3, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 136–142

HOMONYMS AS ITEMS IN VERBAL DISCRIMINATION LEARNING AND TRANSFER

DONALD H. KAUSLER[1] AND RICHARD D. OLSON

Saint Louis University

In Exp. I, verbal discrimination (VD) lists composed of homonym word pairs as wrong (W) and right (R) items were more difficult to learn than a control list containing no homonym word pairs. Moreover, an intrapair W-R homonym condition was less difficult to learn than an interpair W-R condition. Free recall measures of W and R items indicated that the intrapair and interpair conditions differed primarily in the process employed to "tag and collapse" pairs of items for use in the recognition of R items during VD practice. In Exp. II, the W_1-R_1, W_2-R_1' VD transfer paradigm was employed with homonym words providing the R_1-R_1' relationship. Positive transfer, relative to the W_1-R_1, W_2-R_2 control, was found, but only when *S*s were informed that List 2 contained items similar to List 1 items.

The use of homonyms as verbal discrimination (VD) items presents an intriguing test of frequency theory (Ekstrand, Wallace, & Underwood, 1966) via an indirect manipulation of response frequency. Two experiments are reported here, the first relating homonyms to single-list VD learning and the second to VD transfer.

EXPERIMENT I

A VD list composed of homonyms should be more difficult to learn than a list devoid of homonym content, provided the two members of a homonym pair serve separately as wrong (W) and right (R) items within the list. For example, the words *peace* and *piece* may each appear within the list, with *peace* serving as a R item and *piece* as a W item. Lists containing such homonym pairs may be constructed with the homonyms entering into either intrapair or interpair relationships. In an intrapair list, words like *peace* and *piece* would be paired together with *peace* as the R item and *piece* as the W item. In an interpair list, *peace* would be the R item for a pair in which a nonhomonym to *peace* (e.g., *scene*) would be the paired W item. In turn, *piece* would be the W item of a different pair and *seen* would be the R item of still another pair, etc. Verbalization of one member of the pair as a R item is expected to add frequency units to its counterpart present as a W item either within the same pair (the intrapair condition) or within another pair elsewhere in the list (the interpair condition). The increase in absolute response frequency to W items should make intrapair discriminations between W and R items considerably more difficult than in a nonhomonym control condition—to the extent that such discriminations are based on differential response frequencies. In the present experiment, rehearsal responses are presumed to be the primary source of adding frequency units to W items in the homonym conditions. To the extent that *S*'s rehearsal of a given R item, such as *peace*, consists of covert pronunciation of the word, response frequency should generalize to the phonetically identical W item, such as *piece*. The effects of overt pronunciation responses as an additional source of generalization to W items were minimized in the present study by having *S* indicate his selections of R items during anticipation intervals in terms of overt "left"-"right" spatial position responses, rather than verbalizations of specific word choices. This procedure avoided the confusion that would have otherwise occurred in the intrapair homonym condition.

Frequency theory per se does not predict a difference in acquisition rates between intrapair and interpair homonym conditions. However, several recent studies (e.g., Fulkerson & Kausler, 1969; Kausler & Boka,

---

[1] Requests for reprints should be sent to Donald H. Kausler, Department of Psychology, Saint Louis University, St. Louis, Missouri 63103.

1968) suggest that additional processes may provide discriminative cues that compensate, at least partially, for decrements in the proficiency of differential response frequency cues. Especially relevant is the study by Fulkerson and Kausler, in which increments in total frequency to W items were generated by employing bidirectional word associates as W and R items. Intrapair and interpair conditions comparable in form to those of the present study were contrasted with a control condition receiving unrelated words as W and R items. Only the interpair condition was found to be significantly inferior to the control condition. A free recall trial following VD practice provided evidence for a compensatory process which is more effective in the intrapair condition than in the interpair condition. This process consists of *S*'s tagging of the two members of bidirectional associates in terms of "wrong" and "right" labels or features which are then stored in memory and retrieved as cues for subsequent W-R discriminations (Barch, Lippman, & Whalen, 1967). If a similar process operates when homonyms rather than word associates are employed, the interpair condition should again be more difficult than the intrapair condition. Consequently, a free recall task following VD practice was given in the present study as a means of providing further evidence for such compensatory processes.

Method

Subjects and conditions.—The *S*s were 36 undergraduates who were assigned in randomized blocks to the three groups ($n = 12$). All of the *S*s were naive with respect to prior participation in verbal learning experiments. The three groups represented two experimental conditions, Groups E1 and E2, for which the homonyms entered into intrapair and interpair relationships, respectively, and a control condition, Group C, which received unrelated words as W and R items.

Lists.—The homonyms were words of A and AA frequency (Thorndike & Lorge, 1944) that are listed in a reference book on homonyms (Kirtland, 1968). Twenty-four pairs of homonyms were selected as an initial pool. From this pool two different lists, Forms A and B, of 12 pairs each were constructed for Group E1 by random selection from the initial pool. One member of each homonym pair was randomly designated the W item and the other member the R item. Forms A and B of a 12-pair list were prepared for Group E2 by randomly re-pairing the homonym members of Forms A and B for Group E1, with the W and R items retaining the same functions they had for Group E1. Forms A and B of a 12-pair list were prepared for Group C by retaining the R items of the experimental lists and replacing the W items. The replacements were words of A and AA frequency selected from the Palermo and Jenkins (1964) stimulus words. These words ranged in length from two to seven letters (as did the homonyms), and they fell into approximately the same distribution of word classes (verbs, prepositions, etc.) as did the homonym words. The use of two forms, A and B, permitted greater generalization of the findings, as well as greater control of idiosyncratic associations within sets of words. List form (A vs. B) was found to have a negligible effect, both as a main effect and as a source of interaction with the E and C conditions, and the two forms were collapsed for purposes of the analyses reported in the Results and Discussion section.

It must be noted that the use of homonyms as a task variable introduces two problems with obdurate methodological solutions. The first problem occurs via the fact that the preponderance of homonym pairs consists of graphically similar words, especially in the sense that the members of pairs characteristically have the same initial letter. Evidence for this bias is apparent from a sample of 103 pairs of homonyms generated by randomly selecting a pair from each page of the reference source (Kirtland, 1968). Of the 103 pairs, 14.6% consist of pairs with different initial letters. Consequently, an intrapair homonym list constructed from a representative sample of homonyms will contain a vast majority of pairs with graphically similar W and R items. On the other hand, both interpair homonym lists that are constructed by random re-pairing of homonyms and control lists that are constructed by random assignments of replacement words are likely to have pairs with graphically disimilar W and R items. The use of initial letters as a cue for intrapair discriminations is an unexplored area in VD learning, and the likelihood of its involvement as a confounding factor in homonym research is unknown. One solution to the problem consists of restricting the homonym pool and the replacement pool to words that have the same initial letter. This approach, however, introduces several complicating factors. Generalization of the effects of the experimental variable is necessarily limited to a unique subset of homonyms. Moreover, the homonyms included in such a selection procedure would necessarily range greatly in word frequency in order to obtain a sufficient number for list construction, and homogeneity of list content with regard to the frequency dimension was deemed essential in the present study. For these reasons, an alternative approach was taken. The initial pool of homonym pairs was permitted to contain both identical and dissimilar initial-letter words, with the latter being in the minority (6 pairs out of the 24). The increment in the proportion of dissimilar pairs over the proportion indicated by the prior sample (25% vs. 14.6%) permits a more adequate postexperimental analysis of the effects of initial-letter similarity. For this analysis, pairs in the intralist homonym condition were dichotomized into two categories, identical letters ($n = 18$ pairs) and dissimilar letters ($n = 6$ pairs), and mean error rates to criterion were determined separately for each category. The difference in mean errors per item, 19.0 and 20.5 for identical and dissimilar categories, respectively, was slight, and initial letter selection seems an unlikely major source of confounding in the present study.

The second methodological problem stems from the associative relatedness of some homonym word pairs. Consequently, intrapair W-R discriminations for both Groups E1 and E2 are likely to involve implicit associative responses (IARs), with the extent of the involvement and the overall effect on W-R discriminations being a function of the preexperimental associative strength. A search of word association norms (Bilodeau & Howell, 1965; Bousfield, Cohen, Whitmarsh, & Kincaid, 1961; Palermo & Jenkins, 1964) yielded normative data for 15 of the 24 homonym pairs included in the present study. Of this 15 pairs, 7 pairs consisted of associatively unrelated words (0% associative strength). Of the remaining 8 pairs, the maximum associative strength was 2.5%, and the mean associative strength was 1.0%. A similar analysis of the lists for Group C reveals that approximately half of the pairs contained words with some degree of associative relatedness to words elsewhere in the list, either within the same pairing or across different pairings. The maximum associative strength was 7.5%, and the mean was again 1.0%. The *E*s conducted a myriad of postexperimental analyses contrasting error rates and free recall scores for associatively related pairs with associatively unrelated pairs, classified in terms of various intrapair and interpair contingencies, for Groups E1, E2, and C considered separately. In each analysis the effect attributable to associative relatedness was slight, and the uncontrolled presence of associatively related words seems an unlikely major source of confounding in the present study.

Procedure.—The anticipation method was used, with a 2:2-sec. rate of presentation and a 4-sec. intertrial interval. Four different random orders of presentation were employed for each list as a control for serial learning effects. The W and R items of a single pair were typed in juxtaposition and exposed via a Lafayette memory drum. First the pair appeared, and *S* indicated his choice of the R item by identifying its spatial position (left or right); then the pair appeared again with the R item underlined as the means of providing informative feedback. For each pair of items the R item occurred on the spatial right in two of the random orders of presentation and on the spatial left in the other two orders. For any given list there were two pairs representing each of the six possible spatial sequences across orders (right, right, left, left; right, left, right, left; etc.) For each order of presentation half of the R items were spatially on the right, and half on the left. Thus the use of spatial cues for intrapair discriminations was minimized as much as possible. The *S*s in the two homonym groups were not informed in advance of the homonym content of the list. However, each *S* did receive two trials on a practice list composed of two pairs selected from the alternate form (i.e., Form B if *S* received Form A, etc.) of the same list condition. Practice on the experimental list continued to a criterion of 2 consecutive perfect trials or to a maximum of 20 trials. The *S*s were required to make a selection for each pair on each trial, including the first.

Attainment of criterion was followed by a free recall task without a time limit. For this task *S* was given a sheet of paper containing 24 lines, and he was asked to write in as many of the words contained in the list as he could remember, in whatever order he recalled them. After *S* indicated he was finished, he was asked to go over the list of recalled words and indicate for each word whether the word had been a W word or an R word. If *S* recalled additional words during this identification period, he was permitted to record them at the bottom of the list.

Results and Discussion

All *S*s in Group C attained the criterion within 20 trials; one *S* in Group E1 and five *S*s in Group E2 failed to attain the criterion and were assigned scores of 20 for the trials measure. Means and standard deviations for trials to criterion were 13.25 and 3.97, 17.67 and 3.80, and 8.75 and 2.93 for Groups E1, E2, and C, respectively. The between-group effect was significant, $F(2, 32) = 18.51$, $p < .001$, as were the contrasts between means (Newman-Keuls test) for Groups C and E1, $p < .01$, Groups C and E2, $p < .01$, and Groups E1 and E2, $p < .05$. Means and standard deviations for total errors in attaining the criterion were 40.25 and 21.14, 50.08 and 21.40, and 24.83 and 12.15, for Groups E1, E2, and C, respectively. The between-group effect was again significant, $F(2, 32) = 5.50$, $p < .01$. The Group C vs. Group E2 contrast was again significant (Newman-Keuls test), $p < .01$, and the Group C vs. Group E1 contrast approached significance, $p < .10$. However, the Group E1 vs. Group E2 contrast did not approach significance, $p > .10$. In view of the heterogeneity of variance between the two experimental groups and the control group and the skewed distribution of scores in Group E1, the error data were also analyzed by means of the median test (medians = 31, 47, and 22 for Groups E1, E2, and C, respectively). The between-group effect was significant, χ^2 (2) = 6.00, $p < .05$, as were the contrasts between Groups C and E1 and Groups C and E2, both χ^2s (1) = 4.17, $p < .05$. The Group E1 vs. Group E2 contrast again fell short of significance, χ^2 (1) = 1.50, $p > .10$.

The previously stated results provide convincing support for the hypothesized decrement in VD learning proficiency when homonyms serve as W and R items, either in intrapair or interpair W-R relationships. There is also partial support for the hypothesis that the decrement is greater in interpair lists than in intrapair lists. That is, the interpair condition (Group E2) required more trials to attain criterion than did the intrapair condition (Group E1). A greater number of errors occurred in the interpair condition as well, although the difference in errors for the two homonym conditions did not reach statistical significance. Moreover, the reported inferior performance for Group E2 is most likely a conservative estimate in that a substantially larger number of *S*s in this group failed to attain criterion than in either of the other two groups.

As noted earlier, the free recall measure was included as a means of detecting a compensatory process present in the intrapair condition only that would mitigate the deleterious effects of homonyms as W and R items. The free recall scores reported here are for a lenient scoring system in which W and R items are considered to be recalled correctly regardless of the correctness of their subsequent identification as W and R items. Stringent scores, requiring correct item function identification as well as item recall, were found to be nearly identical with the lenient scores. For W-item recall, means and standard deviations were 8.00 and 1.70, 7.50 and 2.35, and 4.00 and 2.21 for Groups E1, E2, and C, respectively. Comparable means and standard deviations for R-item recall were 8.25 and 1.76, 8.42 and 2.06, and 7.92 and 1.91. A 3 × 2 mixed analysis of variance, with list condition as the between-group variable and type of item (W vs. R) as the within-group variable, yielded significant main effects for list condition, F (2, 33) = 4.71, $p < .05$, and type of item, F (1, 33) = 68.00, $p < .001$, and a significant Condition × Item interaction, F (2, 33) = 30.12, $p < .001$. In view of the significant interaction, the effect of list condition was tested separately for W items and for R items. The between-group was significant for W items, F (2, 33) = 12.81, $p < .001$, but not for R items, F (2, 33) < 1. Mean W recall for Group C was significantly below that of both Group E1 and Group E2, p's < .01, Newman-Keuls test, but the latter two did not differ significantly, $p > .10$. Assuming *S*'s awareness of the list's homonym composition, the superior W-item recall for Groups E1 and E2, given an equal amount of R-item recall, is not surprising. That is, *S*'s recall of a R item should mediate the recall of its homonym as well. The presence of such a process was tested by a probability analysis which determined the proportion of W items recalled that are homonyms of recalled R items. This proportion was .97 for Group E1, but only .74 for Group E2, a difference that is statistically significant by the median test, χ^2 (1) = 6.04, $p < .02$. Thus virtually all of the W items recalled by *S*s in Group E1 were the homonym counterparts of recalled R items, whereas a sizable segment of W items recalled by *S*s in Group E2 were words that did not have their homonym counterparts recalled.

Interpretation of the free recall data in terms of direct effects of homonyms on W and R recall is complicated by the fact that the total number of trials to criterion differed significantly between the three groups, and the fact that recall may, in turn, covary with the number of trials. However, the within-group correlation coefficients between trials and amounts of both W and R recall suggest that trial differences are an unlikely source of recall differences. Only the correlation between trials and W-item recall for Group E2 approached significance, r (10)

$= -.56$, $p < .10$; for all other r's, $p > .10$. If anything, increasing trials in the homonym conditions was accompanied by reduced item recall.

The free recall protocols for *S*s in Groups E1 and E2 were also analyzed by clusters of homonym pairs recalled successively. The mean number of clustered pairs was 7.42 ($SD = 3.13$) for Group E1 and 3.67 ($SD = 3.78$) for Group E2. A median test indicated that the difference between the two distributions was significant, χ^2 $(1) = 4.17$, $p < .05$. Absolute number of recalled pairs was employed in this analysis in that the total number of W and R items recalled was nearly identical for the two groups (16.25 for Group E1 and 15.92 for Group E2), and clustering scores were unlikely to be biased by a differential availability of items qua items. Free recall protocols for Group E2 were also analyzed for clusters of intrapair W and R items recalled successively. The mean was 1.33, a value close to the mean of Group C (1.67) for the same measure. Intrapair clusters for Groups E2 and C were also examined by a relative measure, employing the formula C1/N/2, where C1 is the number of intrapair clusters recalled and N is the total number of W and R items recalled (means of 15.92 and 11.92 for Groups E2 and C, respectively; the disparity in means suggesting the necessity of a relative measure that is unbiased by the total number of items recalled). The mean proportion was .18 for Group E2 and .27 for Group C, a difference that is not statistically significant, χ^2 $(1) = 2.68$, $p > .10$. The present results closely parallel those of Fulkerson and Kausler (1969) with bidirectional word associates as W and R items, and they suggest the operation of a compensatory process under homonym conditions akin to that present under bidirectional associative conditions.

EXPERIMENT II

Under appropriate conditions, homonym relationships between items may be expected to generate a positive effect on VD learning, as well as the negative effect demonstrated in Exp. I. The condition selected for a test of this hypothesis in Exp. II was transfer in the W_1-R_1, W_2-R_1' paradigm (i.e., the R items of List 2 bear a relationship to the R items of List 1, but the W items of List 2 are unrelated to those of List 1). The R_1-R_1' relationship here represents pairs of homonym words. Frequency units accruing to R_1 items during List 1 practice may be expected to transfer to their R_1' counterparts in List 2, thereby enabling *S*s to apply the Rule 1 of frequency theory (always select the alternative with the larger accrued response frequency; Ekstrand et al., 1966) early in List 2 practice. Thus positive transfer is the expected outcome, relative to a control condition receiving a W_2-R_2 list (both W and R items of List 2 are unrelated to the W and R items of List 1) as List 2.

Previous research (Kanak & Dean, 1969; Kausler & Dean, 1967) with the W_2-R_1' paradigm has revealed that the transfer effect is only weakly positive when the R_1-R_1' relationship involves associatively related words. However, *S*s in these studies were not informed in advance of the presence of a relationship between the contents of the two lists. Instructions regarding the presence of a list relationship have been previously demonstrated to be a critical factor in determining the direction and extent of transfer for another VD paradigm, W_1-R_1, W_1-R_2 (identical W items in the two lists, but different R items; cf. Kausler, Fulkerson, & Eschenbrenner, 1967), and instructions may likewise be of great importance in the present paradigm. Consequently, the present study included the presence-absence of instructions as an additional treatment condition.

Method

Subjects.—The *S*s were 36 undergraduates who were assigned in randomized blocks to three groups ($n = 12$). All of the *S*s were naive with respect to prior participation in verbal learning experiments. The three groups represented two experimental conditions, Groups E1 and E2, receiving W_2-R_1' Lists 2, with and without instructions concerning an interlist relationship, and a control condition, Group C, receiving a W_2-R_2 List 2.

Lists and procedure.—All three groups received the same List 2, with paradigmatic variation being accomplished by way of List 1. List 2 was identical to the list learned by Group C of Exp. I (again in two forms, A and B). For Groups E1 and E2, the items of List 1 were the homonym counterparts of the List 2 R_1' items; for Group C, the R_1 items were words selected from the Palermo and Jenkins (1964) norms. The W_1 items of List 1 for all three groups were also words selected from the norms, with the items being randomly assigned to pairings with the appropriate set of R_1 items. As in Exp. I, the words selected from the norms closely matched the homonyms in word length, word frequency (A and AA categories), and word class. In addition, the new words selected for this experiment had minimal associative relatedness (no more than 1 per 1,000) with one another. Two different forms, A and B, were also employed for List 1. However, the effects of list form were again found to be slight, and the two forms were collapsed for the analyses reported subsequently.

With one exception, the procedure for presenting each list was identical to that of Exp. I. The exception was the return to the conventional procedure of having *S* verbalize his selections of R items during the anticipation interval. List 1 was practiced to a criterion of two consecutive perfect trials. Following attainment of the criterion, all *S*s were informed that they would now have a second list to learn. In addition, *S*s in Group E1 were further informed that they might notice that some of the words in the new list were very similar to some of the words on the previous list. Five trials were then given on List 2.

Results and Discussion

List 1.—Mean trials to criterion were 8.33, 10.50, and 8.67 for Groups E1, E2, and C, respectively; mean errors in attaining the criterion were 22.41, 28.33, and 21.08, respectively. For both measures the between-group effect was not statistically significant, $F\ (2, 33) = 1.92$, $p > .10$, and $F\ (2, 33) = 1.86$, $p > .10$, thus indicating the comparability of the three groups in List 1 acquisition.

List 2.—Means and standard deviations for total errors across the five trials were 3.00 and 1.85, 13.25 and 7.37, and 12.62 and 2.66 for Groups E1, E2, and C, respectively. Although the between-group effect was significant, $F\ (2, 33) = 16.34$, $p < .001$, Newman-Keuls tests revealed that only the Group E1–Group C and Group E1–Group E2 contrasts were significant, p's $< .01$. Because of the large difference in variability between the groups on error scores, the data were also analyzed by median tests. The results of these analyses confirm the pattern reported previously. The between-group effect was significant, $\chi^2\ (2) = 25.37$, $p < .001$, as were Group E1–Group C and Group E1–Group E2 comparisons, χ^2s $(1) = 13.50$ and 6.04, p's $< .001$ and $< .02$, respectively. However, the Group E2–Group C comparison again fell short of significance, $\chi^2\ (1) < 1$. Finally, it may be noted that 11 *S*s in Group E1 reached a criterion of two consecutive perfect trials within the block of five trials, as compared to 1 *S* in Group E2 and 2 *S*s in Group C.

Discussion

Although not explicitly stated in the frequency theory of Ekstrand et al. (1966), the theory implies that the differential response frequency to intrapair W and R items produced by verbalizations during practice serves as the functional stimulus for the instrumental response of selecting the right item. This functional stimulus is unreliable when homonyms form the items, and the theory clearly predicts the decrement in the learning of homonym lists found in Exp. I. The present results suggest that *S* is likely to abandon the use of differential response frequency as a functional stimulus and is likely to seek instead a reliable compensatory cue as a functional stimulus. A likely compensatory cue is provided by the differential set of IARs occurring to a pair of homonyms. That is, a given homonym R word is tagged during the feedback exposure by some salient IAR which is stored in memory and then retrieved during subsequent anticipation intervals as the functional stimulus for selecting that given word. The present results, like the earlier results of the Fulkerson and Kausler (1969) study, strongly suggest that the tagging process is accomplished more readily when *S* is required to make simultaneous discriminations (the intrapair condition) between related words rather than successive discriminations (the interpair condition). That is, the discriminability of IARs to homonym words is more apparent when the homonyms are presented in juxtaposition than when they are presented successively with other words intervening.

Positive transfer via homonyms, and the resultant use of Rule 1 of frequency theory (Ekstrand et al., 1966) for R-item selection on List 2, occurred in Exp. II only when prior information was given that alerted *S* to expect an interlist relationship. If anything, transfer without this expectancy was slightly negative. Interlist effects of homonyms are seemingly less pronounced than are intralist effects. The temporal separation of homonym pairs from List 1 to List 2 is likely to prevent *S*'s discovery of the relationship unless he is actively searching for a similarity dimension. The direct transfer of verbalization responses that have accrued during List 1 practice is an unlikely mechanism in the W_1-R_1, W_2R_1' paradigm, just as the direct transfer of IARs is an unlikely mechanism in the same paradigm when word associates provide the R_1-R_1' relationship (Kanak & Dean, 1969). This finding may be contrasted with the pronounced positive transfer found in the W_1-R_1, W_2-R_1 paradigm even when *S* is not instructed about the carry over of R items from List 1 (Eschenbrenner & Kausler, 1968). The different set of IARs occurring to the R items in Lists 1 and 2 when homonyms provide the source for transfer of verbalization responses is likely to be the inhibitory process that prevents *S*'s use of Rule 1 at the onset of List 2 practice.

REFERENCES

BARCH, A. M., LIPPMAN, M., & WHALEN, S. Intra-pair relations in verbal-discrimination learning: A paradoxical finding. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1967.

BILODEAU, E. A., & HOWELL, D. C. *Free association norms.* Washington, D. C.: United States Government Printing Office, 1965.

BOUSFIELD, W. A., COHEN, B. H., WHITMARSH, G. A., & KINCAID, W. D. The Connecticut free association norms. (Tech. Rep. No. 35) Storrs, Conn.: University of Connecticut, 1961.

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review,* 1966, 73, 566–578.

ESCHENBRENNER, A. J., JR., & KAUSLER, D. H. Unlearning of List 1 wrong items in verbal discrimination transfer. *Journal of Experimental Psychology,* 1968, **78**, 696–698.

FULKERSON, F. E., & KAUSLER, D. H. Effects of intra-pair and inter-pair bidirectional associates on verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 307–310.

KANAK, N. J., & DEAN, M. F. Transfer mechanisms in verbal discrimination. *Journal of Experimental Psychology,* 1969, **79**, 300-307.

KAUSLER, D. H., & BOKA, J. A. Effects of double functioning on verbal-discrimination learning. *Journal of Experimental Psychology,* 1968, **76**, 559–567.

KAUSLER, D. H., & DEAN, M. G. Direct and mediated transfer for right and wrong responses in verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 672–674.

KAUSLER, D. H., FULKERSON, F. E., & ESCHENBRENNER, A. J., JR. Unlearning of List 1 right items in verbal-discrimination transfer. *Journal of Experimental Psychology,* 1967, **75**, 379–385.

KIRTLAND, E. *Write it right: A handbook of homonyms.* New York: Golden Press, 1968.

PALERMO, D. S., & JENKINS, J. J. *Word association norms: Grade school through college.* Minneapolis: University of Minnesota Press, 1964.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received February 5, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 143–147

DIMENSIONAL SIMILARITY IN CONCEPT IDENTIFICATION AND EXTRADIMENSIONAL SHIFTS

THOMAS O. NELSON[1]

University of Illinois

Ninety-six undergraduates made extradimensional shifts when the relevant preshift and postshift dimensions were either similar or dissimilar. The similar dimensions were: (*a*) height and width (dimensions of size) or (*b*) hue and brightness (dimensions of color). The dissimilar dimensions consisted of one size dimension and one color dimension. Both original learning and extradimensional shifting were easier when the relevant dimensions were dissimilar rather than similar; dimensional similarity accounted for 13.6% of the total variance in preshift performance and 30.8% of the total variance in postshift performance. The increased difficulty of a similar-dimensions shift was attributed to the fusion and subsequent confusion of the relevant dimensions.

The present study had two aims: (*a*) to determine how the similarity of the varying preshift dimensions affects initial concept identification and (*b*) to determine how the similarity between the relevant preshift dimension and the relevant postshift dimension affects extradimensional, or nonreversal, shifting.

There are, of course, many bases for defining "similarity." For the present study, two dimensions are considered to be similar when they themselves can be subsumed under a single superordinate concept or dimension (e.g., *height* and *width* can be subsumed under *size*). Correspondingly, two dimensions are said to be dissimilar when a single superordinate concept is not readily available (e.g., *height* and *brightness* are not easily encompassed by a single superordinate concept).

A primary task in concept identification is discrimination, and it is well known that discrimination becomes more difficult as the stimuli to be discriminated become more similar (Hull, 1943). Therefore, one would expect that if the present notion of dimensional similarity abides by the usual rules, concept identification during original learning should be easier when the varying dimensions (i.e., potentially relevant dimensions) are dissimilar rather than similar.

Also, the similarity between the relevant preshift dimension and the relevant postshift dimension may affect the ease of an extradimensional shift (EDS). Consider the following "fusion-confusion" hypothesis and assume that the EDS is between similar relevant preshift and postshift dimensions, such as *height* and *width*. If these two dimensions are fused into the superordinate *size*, then *size* is partially relevant (to the degree that *height* is relevant) during preshift. Following the shift, *size* becomes partially irrelevant (to the degree that *height* becomes irrelevant). However, if *width* is to become the relevant postshift dimension, then *size* must also be partially relevant *at the same time* as it is partially irrelevant. This anomaly should be confusing to *S* until he reaches the point where he effectively separates *size* into *width* and *height*. Notice that *S* can avoid this confusion if instead of selecting *width* for testing, he selects a dissimilar dimension such as *hue*. The fusion-confusion hypothesis predicts that, other things being equal, *S* will select dimensions for testing that are least similar to the relevant preshift dimension. Only after exhausting all dissimilar dimensions should similar dimensions be selected. Also, if the confusion is great enough, *S* may formulate unnecessarily complex (e.g., conjunctive) hypotheses to test. Thus, the fusion-confusion hypothesis implies that a dissimilar-dimensions EDS should be easier than a similar-dimensions EDS.

---

[1] Appreciation is expressed to William H. Batchelder, William E. Kappauf, Thomas Trabasso, and Joseph L. Wolff for their generous suggestions and clarifications.

Requests for reprints should be sent to Thomas O. Nelson, Department of Psychology, University of Illinois, Urbana, Illinois 61801.

METHOD

Stimuli.—Each stimulus card was white and had an inverted T in its center. The bars making up the inverted T were .75 in. wide. The inverted T could vary on the following five binary dimensions: height of the vertical bar (HT), width of the horizontal bar (WD), hue (HU), brightness (BR), and presence or absence of a .5-in. white dot (a dimension that was always irrelevant). The values for the four relevant dimensions were: for HT, *tall* = 4 in. or *short* = 2 in.; for WD, *wide* = 4 in. or *narrow* = 2 in.; for HU, *red* or *brown*; for BR, *light* or *dark*. The stimulus cards were grouped into six sets. Each card set varied on only three dimensions, two of which were relevant and the third was the irrelevant dot dimension (all other dimensions were held constant). Two of these card sets had similar varying dimensions: (*a*) HT and WD (both are dimensions of size) and (*b*) HU and BR (both are dimensions of color). The remaining four sets had dissimilar varying dimensions: (*a*) HT and HU, (*b*) HT and BR, (*c*) WD and HU, and (*d*) WD and BR. Since each of the three varying dimensions for a given card set was binary, there were $2^3 = 8$ cards in each set.

Subjects and design.—The *S*s were 96 undergraduates from the University of Illinois whose participation fulfilled a course requirement. The design was hierarchical, having the following four variables: (*a*) dimensional similarity, two levels (similar vs. dissimilar); (*b*) pairs of varying dimensions (i.e., card sets), two or four levels nested within each type of dimensional similarity; (*c*) dimensional shift directions, two levels nested within each pair of varying dimensions; and (*d*) positive preshift and postshift values, four levels nested within each dimensional shift direction. The design was replicated twice. Thus, e.g., for two *S*s the positive preshift and postshift values were *tall* → *wide*, nested within *HT* → *WD*, nested within the varying HT and WD card set, nested within the category of similar relevant dimensions. The design was completely symmetrical except that only two groups had similar varying dimensions whereas four groups had dissimilar varying dimensions, which corresponded to the card sets mentioned previously. Thus 32 *S*s made similar-dimensions EDSs, while 64 *S*s made dissimilar-dimensions EDSs. Each *S* saw only one of the six sets of stimulus cards, depending on which pair of varying dimensions he had.

Procedure.—The *S*s were run individually, and each *S* received the same standard instructions (from Buss, 1956). The presentation order of

TABLE 1

MEAN NUMBER OF LOG ERRORS DURING PRESHIFT

Measure	Varying dimensions											
	Similar				Dissimilar							
	HT and WD		HU and BR		HT and HU		HT and BR		WD and HU		WD and BR	
	HT → WD	WD → HT	HU → BR	BR → HU	HT → HU	HU → HT	HT → BR	BR → HT	WD → HU	HU → WD	WD → BR	BR → WD
$\bar{X}$ for dimensional shift direction	1.19	1.71	1.23	1.68	.93	.99	1.16	1.05	1.31	1.28	1.11	.69
$\bar{X}$ for dimension pair	1.45		1.45		.96		1.10		1.30		.90	
$\bar{X}$ for dimensional similarity type	1.45				1.07							

Note.—X = $\log_e(Y + 1)$. For *S*s nested within positive preshift and postshift values, MS = .338; for positive preshift and postshift values nested within dimensional shift direction, MS = .187.

positive and negative instances of the concept was the same for all *S*s and followed a Gellerman (1933) sequence. The first two stimulus cards were complements of each other, insuring that *S* had seen all values of each dimension by the end of the second trial. The first stimulus card after the EDS was chosen so that an error would occur if *S* responded in accord with his preshift hypothesis. The cards were shown from behind a 16×28 in. vertical fiberboard screen, which had a 6×6 in. opening in its base that could be blocked by a guillotine-type door on *E*'s side of the screen.

For both the preshift and postshift phases, *S* was required to reach a criterion of eight successive correct responses. Any *S* who failed to reach the preshift criterion within 50 trials was considered to be a nonlearner and was replaced. (This occurred for only five *S*s, two in the *wide* → *short* group, two in the *narrow* → *short* group, and one in the *brown* → *wide* group.)

RESULTS

Since the number of errors essentially paralleled the number of trials to criterion ($r = .952$), only the error data are reported. For both the preshift and the postshift error data, the means and variances were correlated. Therefore, logarithmic transformations were applied prior to the analyses.

Preshift log errors were analyzed in a four-level hierarchical analysis of variance corresponding to the design discussed previously. The positive preshift and postshift values nested within dimensional shift directions (e.g., the preshift means for *tall* → *wide*, *short* → *wide*, *tall* → *narrow*, and *short* → *narrow*; all are nested within HT → WD) were not significantly different, F (36, 48) = .55. The dimensional shift directions nested within dimension pairs (e.g., the preshift means for HT → WD and WD → HT; both are nested within HT and WD) were not significantly different at conventional significance levels, F (6, 36) = 2.36, $.05 < p < .10$. However, this variable was close enough to significance to warrant reporting the mean number of preshift log errors for each dimensional shift direction, which is shown in Table 1. The dimension pairs nested within dimensional similarity (e.g., the preshift means for HT and WD, and HU and BR; both are nested within the similar-varying-dimensions category) were not significantly different, F (4, 6) = .83. Concerning the variable of major interest, preshift concept identification was more difficult when the relevant dimension was similar, rather than dissimilar, to an irrelevant varying dimension, F (1, 4) = 8.58, $p < .05$; dimensional similarity accounted for 13.6% of the total variance in preshift performance.

The significant differences in preshift performance could have been due to either dimensional similarity or to inherent *S* differences that may have survived the random assignment of *S*s to conditions. If inherent *S* differences were systematically biasing the results, a high correlation between *S*s' preshift and postshift performance would be expected. However, the correlation between preshift and postshift log errors showed no evidence for inherent *S* differences as an explanation for the results ($r =$ .045). Also, to the degree that this correlation is near zero, the preshift and postshift analyses may be regarded as yielding independent information.

The postshift log errors were analyzed in the same manner as the preshift data. The mean number of postshift log errors for each group, down to the level of dimensional shift direction, is shown in Table 2. There were no significant differences between the positive preshift and postshift values nested within dimensional shift directions, F (36, 48) = 1.17. Differences between dimensional shift directions nested within dimension pairs were also nonsignificant, F (6, 36) = .40, as were the differences between dimension pairs nested within dimensional similarity, F (4, 6) = .39. The variable of primary interest, dimensional similarity, was highly significant, F (1, 4) = 128.51, $p <$ .001; the similar-dimensions EDSs were more difficult than the dissimilar-dimensions EDSs. The dimensional similarity variable accounted for 30.8% of the total variance in postshift performance.

DISCUSSION

The advantage of dissimilar dimensions over similar dimensions during the preshift phase suggests that dimensional similarity has a significant effect on initial concept identification. Because similar dimensions may be psychologically fused into a single superordinate concept, *S* may have trouble differentiating them when he is formulating hypotheses to test (cf. the fusion-confusion hypothesis discussed previously). There is no such dimensional fusion, and therefore no differentiation problem, when the varying dimensions are dissimilar.

The fusion-confusion hypothesis was inductively supported during the postshift phase; similar-dimensions EDSs were more difficult than dissimilar-dimensions EDSs. Interestingly enough, postexperimental interviews of the similar-dimensions *S*s (e.g., HT→WD) yielded reports such as, "I wasn't sure what was happening because at first the *big* one was the concept but later it was the wide one." In this case, the superordinate *size,* in place of the actual relevant preshift dimension *height,* may have given rise to confusion during the EDS (cf. Kendler, Kendler, & Sanders, 1967, p. 121). When the relevant preshift and postshift dimensions have a common superordinate, *S* must eventually abandon the superordinate level and instead identify the concept at the subordinate level; partitioning the superordinate into its subordinate components is prerequisite for a similar-dimensions EDS. Contrariwise, in the dissimilar-dimensions EDS (e.g., HT → HU) either the superordinate concepts (e.g., *size* and *color*), the subordinate concepts (e.g., *height* and *hue*) or combinations of the two (e.g., *size* and *hue*) can be used effectively because there is no need to discriminate below the superordinate level.

Notice that the fusion-confusion hypothesis makes a prediction that is opposite to one extrapolated from the conventional mediation model of conceptual shifting. For example, Kendler and Kendler (1962) report that a conceptual shift is easiest when the positive preshift and postshift *values* are mediated by a single dimension (e.g., the shift *tall* → *short* is easier than the shift *tall* → *red*). Therefore one might expect an EDS to be easiest when the relevant preshift and postshift *dimensions* are themselves "mediated" by a single superordinate concept. Such an extrapolation implies that a similar-dimensions EDS would be easier than a dissimilar-dimensions EDS, which clearly was not the case in the present study.

There is no necessity to interpret the present findings solely within the framework of linguistically mediated shift behavior. They would seem also to fit quite well within an attentional model such as that proposed by Trabasso and Bower (1968, see especially Ch. 6). Then the explanation for the advantage of dissimilar dimensions over similar dimensions might be expressed in terms of cue sali-

TABLE 2

MEAN NUMBER OF LOG ERRORS DURING POSTSHIFT

Measure	Relevant preshift and postshift dimensions											
	Similar				Dissimilar							
	HT and WD		HU and BR		HT and HU		HT and BR		WD and HU		WD and BR	
	HT → WD	WD → HT	HU → BR	BR → HU	HT → HU	HU → HT	HT → BR	BR → HT	WD → HU	HU → WD	WD → BR	BR → WD
$\bar{X}$ for dimensional shift direction	2.04	2.07	1.90	2.43	1.50	1.48	1.25	1.48	1.32	1.33	1.36	1.39
$\bar{X}$ for dimension pair	2.05		2.17		1.49		1.37		1.33		1.38	
$\bar{X}$ for dimensional similarity type	2.11				1.39							

Note.—X = log$_e$Y. For *S*s nested within positive preshift and postshift values, *MS* = .473; for positive preshift and postshift values nested within dimensional shift direction, *MS* = .552.

ence, particularly as it affects the attending response of *S* in the "searching for a dimension to test" stage. An overgeneralized attending response would be disadvantageous in the similar-dimensions situation.

Smiley and Weir (1966) and Downing (1968) have shown that knowledge of *S*s' idiosyncratic dimensional preferences is useful for predicting concept identification success. Such an R-R relation could be coupled with the present S-R relation in order to maximize the overall predictability of concept identification behavior.

Researchers who investigate concept identification and EDSs should be wary of two pitfalls. First, multidimensional constructs such as size or color should not be confused with unidimensional concepts. Second, the various dimensions under investigation should not always be assumed to be psychologically orthogonal; even worse, the possibility may be overlooked that the lack of orthogonality, as in the similar-dimensions situation, can be a determinant of concept identification behavior.

REFERENCES

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *Journal of Experimental Psychology,* 1956, 52, 162–166.

Downing, B. D. Salience and learning rate in concept identification. *Psychonomic Science,* 1968, 10, 73–74.

Gellerman, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *Journal of Genetic Psychology,* 1933, 42, 206–208.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychological Review,* 1962, 69, 1–17.

Kendler, H. H., Kendler, T. S., & Sanders, J. Reversal and partial reversal shifts with verbal material. *Journal of Verbal Learning and Verbal Behavior,* 1967, 6, 117–127.

Smiley, S. S., & Weir, M. W. Role of dimensional dominance in reversal and nonreversal shift behavior. *Journal of Experimental Child Psychology,* 1966, 4, 296–307.

Trabasso, T., & Bower, G. H. *Attention in learning: Theory and research.* New York: Wiley, 1968.

(Received February 7, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 148–155

ROLE OF FEEDBACK STIMULI IN RESPONSE DISCRIMINATION AND DIFFERENTIATION[1]

EDWARD J. RICKERT[2]

University of New Mexico

Two roles of feedback stimuli were assessed by comparing the behavioral outcomes of feedback against the transfer of discrimination property of external stimuli. Three groups of four rats were assigned to two transfer conditions—gradual and abrupt—and a hard condition in each of two experiments. The *S*s using feedback of previous responses as cues for the selection of other responses showed that a difficult discrimination among feedback with respect to ensuing behavior was facilitated by initial training on an easy response discrimination, and *S*s using feedback from concurrent responding to mediate selection among those responses showed that learning a difficult differentiation was facilitated by learning an easier differentiation. The findings: (*a*) show feedback stimuli are sufficient for the selection of ensuing and concurrent behavior and (*b*) indicate that feedback share a common property with external stimuli.

Response-produced-cue theory commonly ascribes two roles to feedback stimuli. One role of these inferred events is to guide ensuing behavior; feedback from previous responses become a part of the total stimulus complex to which the next response is made. The other role, more recently formulated, is to guide concurrent behavior; impending feedback from alternative responses may somehow mediate selection among them. Notwithstanding the early importance assigned to feedback stimuli (e.g., Hull, 1930) and the paramount significance they have recently acquired in learning theories (e.g., Miller, 1963; Mowrer, 1960), the paucity of data concerned with feedback suggests that they lack solid experimental support. Typically, the mediating response has been indirectly indexed by showing that instrumental behavior may be either facilitated or interfered with when a classically conditioned stimulus previously associated with the response sustaining the behavior chain is presented (Estes, 1948; Shapiro & Miller, 1965). Shapiro and Miller questioned the presumptive evidence of identifying the mediational chain in this way because showing that one response (R_1) mediates another (R_2) requires demonstrating that R_1 provides the necessary and sufficient cues for R_2. This condition can be met only if control is exercised over the feedback produced by R_1. In short, a paradigm must permit the same experimental manipulations of feedback that are possible with external stimuli. The purpose of this research was (*a*) to evaluate the two roles assigned to feedback stimuli by devising paradigms which control the temporal sequencing of response-produced stimulation and (*b*) to adduce evidence for response-produced-cue theory by showing that responses mediated by feedback exhibit the same phenomena that are observed when those responses are controlled by external stimuli.

The particular property of external stimuli chosen to render a comparison between external and feedback stimuli was transfer of discrimination, which refers to the finding first reported by Pavlov (1927, pp. 121–123) that learning a difficult discrimination may be facilitated by initial training on an easier discrimination along the same continuum.

The two experiments reported here were performed to determine whether transfer

---

[1] This research is the substance of a dissertation submitted to the Graduate School, University of New Mexico, in partial fulfillment of the requirements for the PhD degree. This investigation was supported by a National Science Foundation grant awarded Frank A. Logan, dissertation chairman. The author expresses his indebtedness to Logan for his guidance and encouragement.

[2] Requests for reprints should be sent to Edward J. Rickert, who is now at Sacramento State College, Department of Psychology, 6000 Jay Street, Sacramento, California 95819.

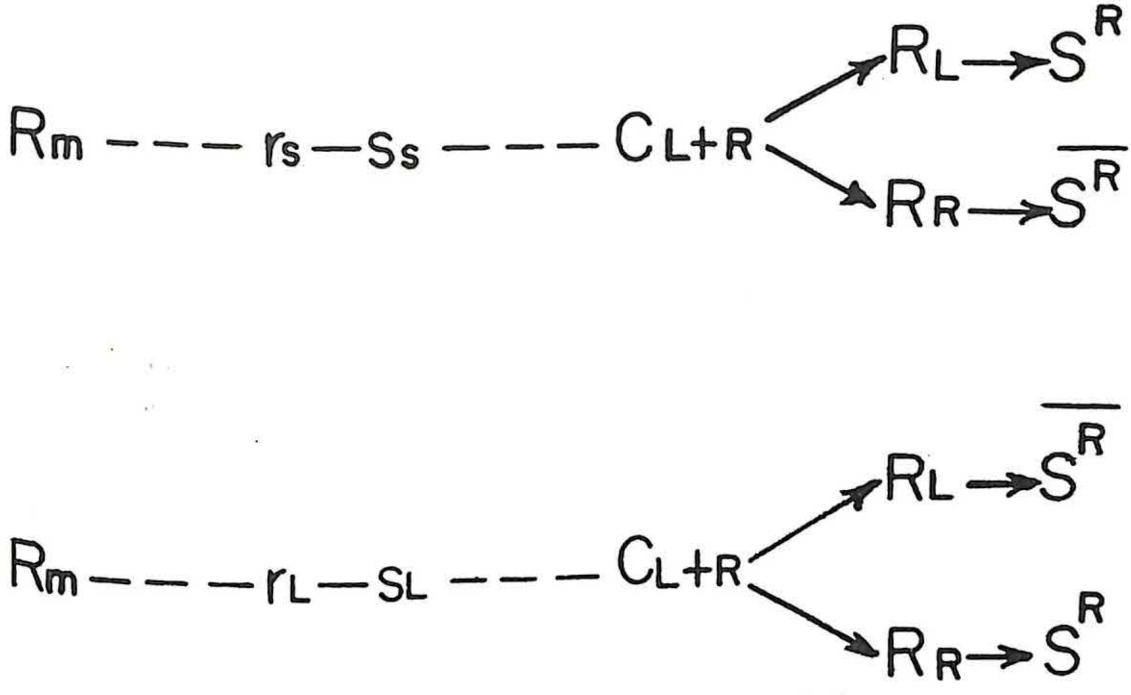


FIG. 1. Schematic representation of response discrimination.

effects comparable to those for external stimuli can be shown for the two roles of feedback indicated previously. The first experiment was concerned with response discrimination; it permitted *S*s to use feedback of previous responses as cues for the selection of other responses. The second experiment, by permitting *S*s to use the impending feedback from quantitatively different responses to mediate selection between those responses was concerned with response differentiation. The distinction between these experiments rests upon the temporal sequencing of potential response-produced stimulation; this distinction is developed more fully in the following section.

Response discrimination and differentiation.—Consider a response chain consisting of R_x followed by R_a and R_b. For the response discrimination procedure, a predetermined duration of R_x causes the onset of external stimuli signaling termination of R_x and selection between R_a and R_b. Which of the latter responses is correct depends on the duration of R_x preceding the signal. The presumed basis for selection between R_a and R_b is then the distinctive feedback resulting from different durations of R_x. Since these feedback cues are available before the choice is made, it may be considered a discrimination based on preceding behavior. In short, response discrimination is when the organism selects what to do next on the basis of what he has already done.

By contrast, response differentiation requires a selection of a duration on either R_a or R_b depending on which of the two is signaled as the correct response. That is to say, one duration is required if R_a is to be reinforced and another duration is required if R_b is to be reinforced. Whereas in response discrimination, external stimuli serve to terminate R_x and enable choice between R_a and R_b on the basis of what has already transpired, in response differentiation, the external stimuli indicate both which response, R_a or R_b, is next to occur and the duration of R_a or R_b necessary for that response to be reinforced. Accordingly, feedback from R_a or R_b is at best available concurrently, and more reasonably, *after* the choice is made. In short, response differentiation is when the organism selects what to do now on the basis of what he is going to do later.

Two distinctions between these situations should be emphasized. First, feedback from R_x as the relevant cue is available to guide selection of ensuing responses in discrimination, whereas it occurs coincident with response selection in differentiation. Second, although the organism controls selection and initiation of response in both situations, response termination is controlled by an external stimulus in discrimination and by the organism in differentiation.

EXPERIMENT I

Transfer of response discrimination.—In order to study the role of feedback stimuli in guiding ensuing behavior, an operant conditioning box containing three bars was used. The middle bar corresponds to R_x in the preceding discussion, while the two side bars correspond to R_a and R_b. The procedure is schematically represented in Fig. 1.

Each trial began with a response to the middle bar (R_m). After a predetermined time period, cue lights (C_L and C_R) appeared above both the left (R_L) and right (R_R) bars. If the duration of R_m preceding the cue lights was short (r_S), the left bar was correct. If the duration of R_m was long (r_L), the right bar was correct. Presumably, the basis for the choice was the differential feedback produced by the response duration on the middle bar (s_S and s_L, in Fig. 1).

The hypothesis based on the notion that feedback stimuli obey the laws of external stimuli was that a difficult response discrimination benefits from initial training on an easy response discrimination. Ease of discrimination in this context was presumed to depend on the size of the difference between the short and long durations of R_m.

Method

Subjects.—The *S*s were 12 female rats of Long-Evans stock from the breeding colony of the University of New Mexico. The *S*s were approximately 100 days old at the time the deprivation cycle was imposed. They were housed individually with water freely available and received a daily maintenance diet of 12 gm/day immediately after being run.

Apparatus.—The experimental chamber was constructed by Lehigh Valley Electronics and measured 10 × 11 × 12 in. The intelligence panel housed three bars with indicator lights located above the left and right bars. The middle bar was mounted directly above a 1½ × 1½ × 2 in. food receptacle into which a single 20-mg. Noyes rat pellet was delivered as reinforcement. The middle bar had a distinctive shape, size, and mounting from the side bars. All events were controlled by commercial programming equipment; responses were recorded on impulse counters.

Procedure.—Prior to assignment to their experimental conditions, *S*s were (*a*) trained to respond differentially to the cue lights associated with the side bars when presented singly, (*b*) required to press the middle bar to produce cue onset randomly over one or the other side bar, and (*c*) shaped to progressively longer durations on the middle bar beginning with .2 sec. and terminating with 1.4 sec. During Stage *c* of pretraining, *S*s were required to depress the middle bar for .2 sec. to produce the cue. On each succeeding day an additional .2-sec. requirement was imposed until *S*s could produce the cue only after depressing the middle bar for 1.4 sec. Each *S* completed 100 trials/day on each duration value. The entire pretraining period required 20 days.

Design.—The design called for the random assignment of four rats to each of three groups: a gradual transfer group (G); an abrupt transfer group (A); and a hard discrimination group (H). The *S*s in Groups G and A were begun on the easy problem: after a response of predetermined duration (either .2 or 1.2 sec.) on the middle bars, onset of the indicator lights occurred over *both* side bars. Since light onset occurred over both bars, the sole basis for deciding which side bar to press presumably resided in the differential feedback produced by the response duration on the middle bar. Each of the two durations was associated with a specific bar, e.g., .2-sec. press durations meant "go right" and 1.2-sec. press durations meant "go left." Durations of .6 and .8 sec. were imposed on Group H from the beginning, again with each duration correlated for any given *S* with one of the side bars. To counterbalance position preferences, two *S*s in each group encountered the opposite correlation between duration and side bar.

After Groups G and A had received 800 trials at 100 trials/day, 50 to each duration, and had apparently achieved asymptote on the easy task, Group G was shifted to an intermediate problem with durations of .4 and 1.0 sec. for an additional 400 trials. Group A continued on the easy task (.2 vs. 1.2 sec.) during this phase.

Both the G and A groups were then shifted to the difficult problem and all three groups received an additional 800 trials on the hard task.

Results

Percentage of choice for successive blocks of 100 trials is shown in Fig. 2. The choice measure was determined by summing the

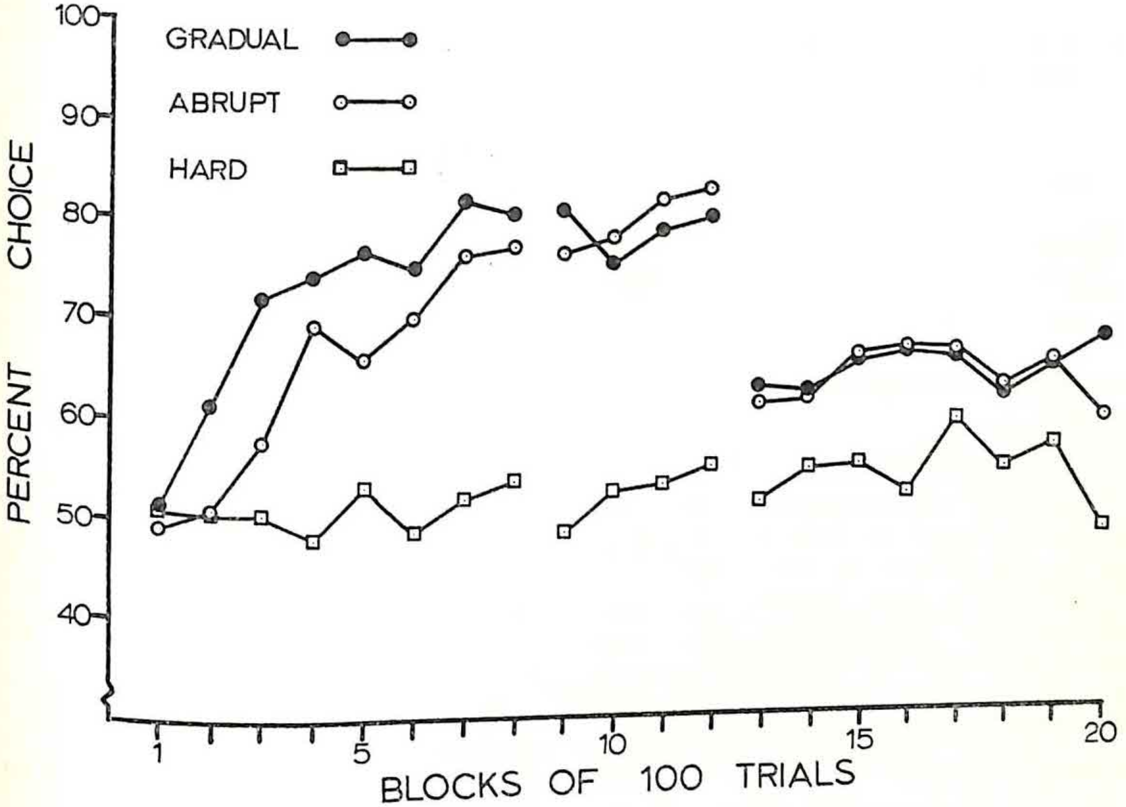


FIG. 2. Response discrimination curves as a function of training conditions. (Gaps in curves represent shifts in condition by one or more transfer groups.)

number of correct responses over both R_a and R_b for each session (100 trials/day), 50 trials to each bar. Since a trial was terminated by a press to either R_a or R_b after cue onset, errors occurred when *S* pressed R_a when R_b was programmed or R_b when R_a was in effect (e.g., 50 − correct R_a = incorrect R_b; 50 − correct R_b = incorrect R_a). The application of Krusal-Wallis analysis of variance by ranks revealed a significant difference between groups over the first eight trial blocks, $H\ (2) = 9.85$, $p < .008$. This indicates that *S*s discriminating between durations of 1.2 and .2 sec. learned that discrimination more rapidly than *S*s required to discriminate between .6 and .8 sec.

The second set of data points (Blocks 9–12) reflects the performance of the groups after Group G was shifted to the intermediate condition, i.e., duration of .4 and 1.0 sec. As may be seen, this condition produced only a slight decline in Group G's performance relative to Group A.

The crucial question of whether initial training on an easier discrimination improves performance on a difficult discrimination above that for Group H is reflected in the set of curves from Blocks 13–20. The apparent superiority of the transferred groups was tested by a Krusal-Wallis analysis of variance on each *S*'s mean overall choice score summed over the last eight trial blocks (see Table 1). The analysis revealed a significant effect among the three groups, $H\ (2) = 5.89$, $p < .05$. Post-hoc comparisons (Mann-Whitney *U* Tests) were made to determine which of the groups differed from one another; the test indicated that the difference between G, A, and H arose from the transfer groups' practice on the easier discrimination, $U\ (8, 4) = 2$, $p < .016$. These comparisons indicate that response-produced stimuli guide ensuing behavior in much the same way as external stimuli in the sense that they share the

TABLE 1

MEAN CHOICE SCORES FOR *S*s IN THE GRADUAL, ABRUPT, AND HARD GROUPS SUMMED OVER LAST EIGHT TRIAL BLOCKS

Cond.	Score			
	*S*1	*S*2	*S*3	*S*4
Gradual	68.2	61.8	59.1	67.5
Abrupt	75.8	56.2	67.5	57.8
Hard	55.6	58.2	45.1	54.4

property of facilitating transfer of discrimination.

EXPERIMENT II

Transfer of response differentiation.—The purpose of Exp. II was to determine whether the learning of a difficult differentiation may be facilitated by initial training in an easier differentiation. If feedback presumably generated concurrent with responding obey the laws of external stimuli, then comparable phenomena should be obtained.

Method

Subjects.—The *S*s were experimentally naive, female rats of Long-Evans stock, ranging in age from 90 to 100 days, obtained from the same source and exposed to the same preexperimental conditions as those in the previous experiment.

Apparatus.—Except for the use of two additional Hunter Decade Timers, the apparatus was the same as that used in the previous experiment.

Procedure.—The *S*s were (*a*) pretrained to respond differentially to the cue lights associated with the side bars when presented singly, (*b*) required to press the middle bar to produce cue onset randomly over one of the other side bars, and (*c*) shaped to progressively longer minimum duration on the side bars beginning with .2 sec. and terminating with 1.4 sec. Each *S* had by the twenty-first day of training completed 100 trials on the 1.4-sec. value.

Design.—As in Exp. I, *S*s were randomly assigned to a G, A, or H group, four *S*s to each group. The *S*s in the G and A groups were placed on the easy differentiation; a press to the middle bar produced the onset of an indicator light over *only* one of the side bars. Light onset served to inform *S* (*a*) which of the two side bars was correct, and (*b*) which duration associated with each side bar to emit before release of that bar (termination of the press within bands of .2–.4 sec. and 1.2–1.4 sec. resulted in reinforcement). In contrast to Exp. I, R_x only served to initiate the trial. Selection of the incorrect bar, the bar not associated with the cue light, had no effect on that trial. Since no external stimulus was available to signal *S* to terminate the response, the presumed basis for his decision resided in the impending feedback available at best coincident with bar release. When durations were imposed on the side bars, only two events were designated as "responses": bar release must have either exceeded the prescribed band or it must have occurred within the band. If *S* released the bar before the timer activated the lower limit of the band, the cue light remained on and that trial remained in force. This procedure attempted to obviate the differential effects of delay of reinforcement. Delay in the present case refers to the delay inherent in the response itself, i.e., the duration of that response. Each of the two bands was associated with a specific bar, e.g., .2–.4 sec. correlated with the right bar and 1.2–1.4 sec. correlated with the left bar. Durations of .6–.8 sec. and .8–1.0 sec. were imposed on Group H from the beginning. Two *S*s in each group encountered the opposite correlation between band and side bars.

After Groups G and A had achieved stability on the easy task (1,400 trials at 100 trials/day, 50 to each bar), Group G was transferred to an intermediate condition with bands of .4–.6 and 1.0–1.2 sec. for an additional 400 trials. During this phase Group A remained on the easy task (.2–.4 vs. 1.2–1.4 sec.) and Group H remained on the difficult task (.6–.8 vs. .8–1.0 sec.).

To determine whether experience on the easier tasks facilitated the learning of a difficult differentiation, Groups G and A were then shifted to the difficult problem and all three groups received an additional 800 trials on the hard task.

Results

The performance of the three differentiation groups over the entire course of training is shown in Fig. 3. This graph shows the total percentage frequency of correct response duration combined for the two responses. It is clear by inspection that all groups improved and hence achieved some degree of response differentiation. It is also apparent in Fig. 3 that the groups did not differ appreciably during the first 14 sessions, $H\ (2) = .19$, $p = ns$. At first glance, this result is surprising since the more widely separated bands should comprise an easier differentiation. Closer inspection of the data, however, suggests that the hard task was indeed more difficult to differentiate, but was an easier performance task. That is to say, *S*s could learn only to differentiate one response duration between 6 and 1.0 sec. and receive intermittent reinforcement on both bars for this larger bandwidth response. This solution would result

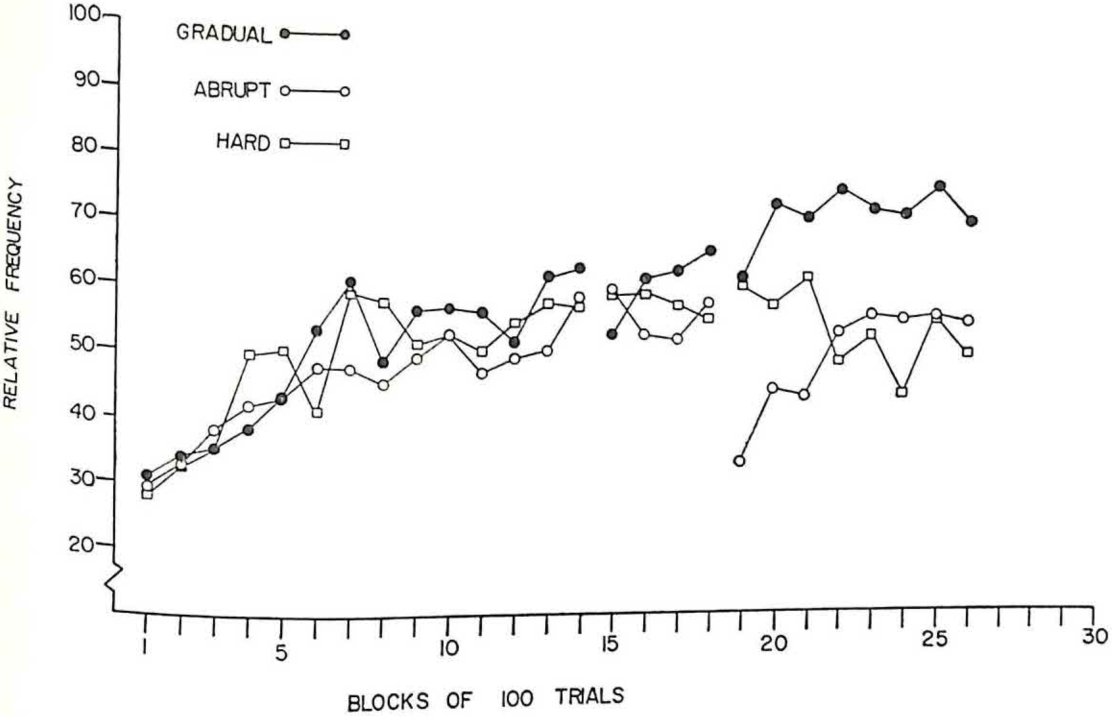


Fig. 3. Response differentiation curves as a function of training conditions. (Gaps in curves represent shifts in condition by one or more transfer groups.)

in about 50% reinforcement when the short band was required and somewhat more than 50% reinforcement when the long band was required, the latter because of the procedure of not terminating a trial when the response duration did not reach the lower limit of the band.

An indication that *S*s in Group H failed to differentiate between the durations may be seen in Table 2 by examining the number of presses made before selecting either a correct or incorrect response, i.e., the presses which failed to terminate the trial. The values in Table 2 represent the mean number of presses to each band before the selection of either a correct or incorrect response. The means were calculated by summing across *S*s within each group over successive trial blocks. Table 2 shows that *S*s in Group H averaged 76.6 presses on the short band and 94.1 on the long band during the first five sessions, and little change ensued throughout training, with 80.8 presses made to the short band as compared to 116.4 on the long band by the last four sessions. Presumably, *S*s adopted a strategy of gradually increasing the duration of successive presses since the procedure required them to reach the lower limit of the band to terminate a trial.

By contrast, inspection of Table 2 for Group G provides clear evidence of differentiating response durations. Although Group G shows a similar pattern of responding with 16.5 presses to the short band and 134.9 to the long band during the first 5 days of training, a shift to the difficult task produced an average of 114.1 presses to the short band and 72.4 presses to the long band even during the last four transfer sessions. The reversal in the number of presses on the two bands indicates that *S*s in Group G emitted response durations more appropriate to the earlier condition. This is to say, durations shorter than those required on the short band and longer than those required on the long band were obtained after the transfer to the difficult task. A similar effect was also observed for Group A after its shift.

A comparison of the performance of Groups G and H in Table 3 in terms of per-

TABLE 2

Mean Presses (Bar Release) on the Two Bands before Selection of Correct or Incorrect Responses Summed over Successive Trial Blocks

Trial blocks	Cond.					
	Gradual		Abrupt		Hard	
	r_S	r_L	r_S	r_L	r_S	r_L
1–5	16.5	134.9	30.1	137.5	76.6	94.1
6–10	31.1	148.5	51.7	163.1	79.6	123.7
11–14	57.2	125.5	50.1	144.1	89.0	126.3
15–18	53.3	63.4	50.1	116.2	91.4	142.6
19–22	108.3	41.8	164.3	60.7	82.5	134.1
23–26	114.1	72.4	149.5	60.9	80.8	116.4

Note.—Subscripts attached to r's denote short (r_S) and long (r_L) durations.

TABLE 3

Mean Percentage of Correct Responses on the Two Bands Summed over Successive Trial Blocks

Trial blocks	Cond.					
	Gradual		Abrupt		Hard	
	r_S	r_L	r_S	r_L	r_S	r_L
1–5	24.4	48.2	33.6	42.6	39.4	41.5
6–10	42.8	69.4	43.4	54.2	50.4	47.8
11–14	46.7	75.9	56.6	46.4	54.0	58.4
15–18	51.4	74.2	61.4	52.6	53.6	59.8
19–22	62.6	82.6	35.6	57.4	58.8	57.4
23–26	70.2	77.6	48.8	63.8	51.9	62.3

Note.—Subscripts attached to r's denote short (r_S) and long (r_L) durations.

centage of correct responses suggests that it was indeed easier for Group G to differentiate the widely separated bands as well as those between more similar durations. During the initial five sessions, Group G achieved 24.4% correct on the short band and 48.2% correct on the long band, a difference presumably resulting from the shaping procedure which initially rewarded relatively long responses. Prior to their first transfer, short-band performance had improved to 46.7% and their long-band performance to 75.9%, an outcome that could not be achieved without clear differentiation of both bands, and their terminal performance on the difficult task showed 70.2% correct on the short band and 77.6% on the long band. In contrast, *S*s in Group H averaged 39.4% correct on the short band and 41.5% correct on the long band over the first five sessions and gradually improved to 51.9% correct on the short band and 62.3% correct on the long band by the last 4 days of training. Only one *S* in Group H did better than 50% correct on both bands, which would provide clear evidence of differentiating between them.

This last result is reflected in Fig. 3 over Blocks 19–26 and addresses the critical issue of whether training on the easy task facilitates learning a more difficult differentiation. A Krusal-Wallis rank-order analysis of variance was performed on the mean percentage of correct responses made by the three groups summed over the eight trial blocks on the difficult task; the difference among groups was significant, $H\ (2) = 8.56,\ p < .016$. Post-hoc comparisons revealed that the significant source of variance may be attributed entirely to Group G: G vs. H, $U\ (4, 4) = 1,\ p < .029$; G vs. A, $U\ (4, 4) = 0,\ p < .014$; while there was no difference between Groups A and H: A vs. H, $U\ (4, 4) = 6,\ p = ns$. This last result, however, may reflect the use of the stringent requirement of including the early trials when Group A suffered a severe transfer decrement, their terminal performance being somewhat above Group H.

Discussion

The results of the two experiments demonstrate that inferred response-produced stimuli obey one of the laws familiar with exteroceptive stimuli: namely, that training on any easy task may facilitate learning a more difficult task. Using response duration as the relevant feedback dimension, response discrimination was defined as selecting ensuing behavior on the basis of available feedback stimuli; training on widely separated durations facilitated learning to discriminate more similar durations. Using the same response dimension, response differentiation was defined as emitting the correct duration on the basis of exteroceptive stimuli; training on widely separated durations facilitated learning to differentiate between more similar durations. These results appear to demonstrate that response-produced stimuli are a sufficient basis for selection of both ensuing and coincident behavior, and the results are thus consistent with the roles ascribed to these inferred events by a variety of contemporary learning theorists (e.g., Miller, 1963; Mowrer, 1960; Sheffield, 1966).

The problem remains as to how to account for the difference in performance between the abrupt transfer groups in the two experiments. One possibility is that since exteroceptive stimuli in Exp. I controlled bar release, they presumably enhanced the distinctiveness between the two durations. This is to say, altering the conditions in Exp. I immediately informed the organism of the change; the lack of any external change in the differentiation experiment meant that *S* continued to emit response-produced stimuli appropriate to an earlier condition for a longer period. The necessity to change behavior and its associated feedback without the support of external stimuli probably accounts for the larger transfer decrements obtained with the differentiation groups. Moreover, since the differential training on Trial Blocks 15–18 in Exp. II required the gradual differentiation groups to produce feedback more similar to those encountered on the difficult task than the abrupt groups, the training enabled generalization mechanisms to exert a more powerful effect on the gradual groups and is probably responsible for the difference between them.

Establishment of feedback as a sufficient condition in directing ensuing and coincident behavior supports a molecular analysis of the response (e.g., Logan & Wagner, 1965; Mowrer, 1960). A molecular analysis contends that the characteristic "feel" of response-produced stimuli feeds back into on-going behavior and directs it in much the same way as external stimuli. The research reported here supports a conception of response which is closed-loop in the sense that an organism selects specific behaviors because it discriminates the stimuli arising from the act of emitting those behaviors. A further implication of this approach is that response differentiation may be subsumed as a subset of stimulus discrimination; the sole distinction between these classes of events is the locus (internal or external) of the cue controlling behavior (Notterman & Mintz, 1965). Nevertheless, some closed-loop formulations posit an identity between the response and its feedback (Mowrer, 1960) and, consequently, cannot account for response initiation. An equally plausible alternative to a "molecular" analysis which obviates the problem of response initiation is a dynamic, linear model. It assumes an organism capable of continuously monitoring its behavior over time so that it decides to continue or terminate responding on the basis of preceding response-produced stimuli. Further, since the results establish that preceding and impending feedback are sufficient for the selection of different responses, the linear model implies that response differentiation is a special case of response discrimination.

REFERENCES

Estes, W. K. Discriminative conditioning: II. Effects of a Pavlovian conditioned stimulus upon a subsequently established operant response. *Journal of Experimental Psychology,* 1948, **38,** 173–177.

Hull, C. L. Knowledge and purpose as habit mechanisms. *Psychological Review,* 1930, **57,** 511–525.

Logan, F. A., & Wagner, A. R. *Reward and punishment.* Boston: Allyn & Bacon, 1965.

Miller, N. E. Some reflections on the law of effect produce a new alternative to drive reduction. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: University of Nebraska, 1963.

Mowrer, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

Notterman, J. M., & Mintz, D. E. *Dynamics of response.* New York: Wiley, 1965.

Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) Oxford: Clarendon Press, 1927.

Shapiro, M. M., & Miller, T. M. On the relationship between conditioned and discriminative stimuli and between instrumental and consummatory responses. In W. P. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965.

Sheffield, F. D. A drive induction theory of reinforcement. In R. N. Haber (Ed.), *Current research in motivation.* New York: Holt, 1966.

(Received February 8, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 156–162

DISCRIMINATION LEARNING AS A FUNCTION OF PRIOR DISCRIMINATION AND NONDIFFERENTIAL TRAINING[1]

KENNETH O. ECK, RICHARD C. NOEL, AND DAVID R. THOMAS[2]

University of Colorado

Four groups of pigeons were given variable interval (VI) training to a vertical (90°) white line in Stage 1 of the experiment. For two of these groups (true discrimination, TD; and no-line control, NLC), a 60° line was alternately presented and responses to it were extinguished; a third group (pseudo discrimination, PD) was equally reinforced for responding to both lines; a fourth group (single-stimulus control, SSC) experienced only the vertical. In Stage 2, all groups were trained to discriminate between two successively presented wavelengths, with a white vertical line superimposed on each (except in the case of NLC). The TD and NLC groups mastered the wavelength discrimination equally rapidly and significantly more rapidly than the PD and SSC groups, which did not differ. At this point, the SSC group was dropped and the other three groups retrained under the original condition (Stage 3). Then, in Stage 4, *S*s learned a brightness discrimination in which, for TD and PD groups only, the vertical line was superimposed on both values. Again the TD and NLC groups performed equally and were significantly superior to the PD group. Thus in two different situations, discrimination training facilitated the subsequent acquisition of a new discrimination independent of the first. In neither case did the presence of the previously positive line angle affect the mastery of the new discrimination, suggesting that "attention" to one stimulus dimension (line angle) does not necessarily detract from attention to others.

Many investigators (e.g., Honig, 1969a, 1969b; Reinhold & Perkins, 1955; Thomas, 1969; Thomas, Freeman, Svinicki, Burr, & Lyons, 1969) have demonstrated that successive discrimination training along one stimulus dimension sharpened gradients of generalization measured along orthogonal stimulus dimensions which were either present but irrelevant during training or were experienced later. This sharpening has been measured relative to both single-stimulus and nondifferential training conditions. Thomas (1969) argued that discrimination training produces a "general attentiveness" or a set-to-discriminate which is not specific to the training dimension, but rather is applied to others as well. A direct implication of this conceptualization is the prediction that discrimination training will facilitate the subsequent acquisition of a second discrimination which is orthogonal to the first. The major purpose of the present study was to test this prediction by first training pigeons to discriminate between two line angles and then noting transfer of training to a discrimination involving two wavelengths. A subsequent test involved a discrimination between two brightness values.

A secondary purpose of the present study was to test the assumption taken from various stimulus selection or selective attention theories (cf. Lawrence, 1950; Mackintosh, 1965; Sutherland, 1964; Trabasso & Bower, 1968) that the more *S* attends to one aspect of a complex stimulus, the less *S* (necessarily) attends to others. This test was accomplished by comparing discrimination performance in the wavelength problem by two groups of *S*s both of which had received identical training to discriminate on the basis of line angle. For the true discrimination (TD) group, the previously positive

---

[1] This research was supported by Research Grant NIH-MH-10427 from the National Institute of Mental Health under the direction of the third author. These results were reported at the meetings of the Psychonomic Society, St. Louis, Missouri, October 1968.

[2] Requests for reprints should be sent to David R. Thomas, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

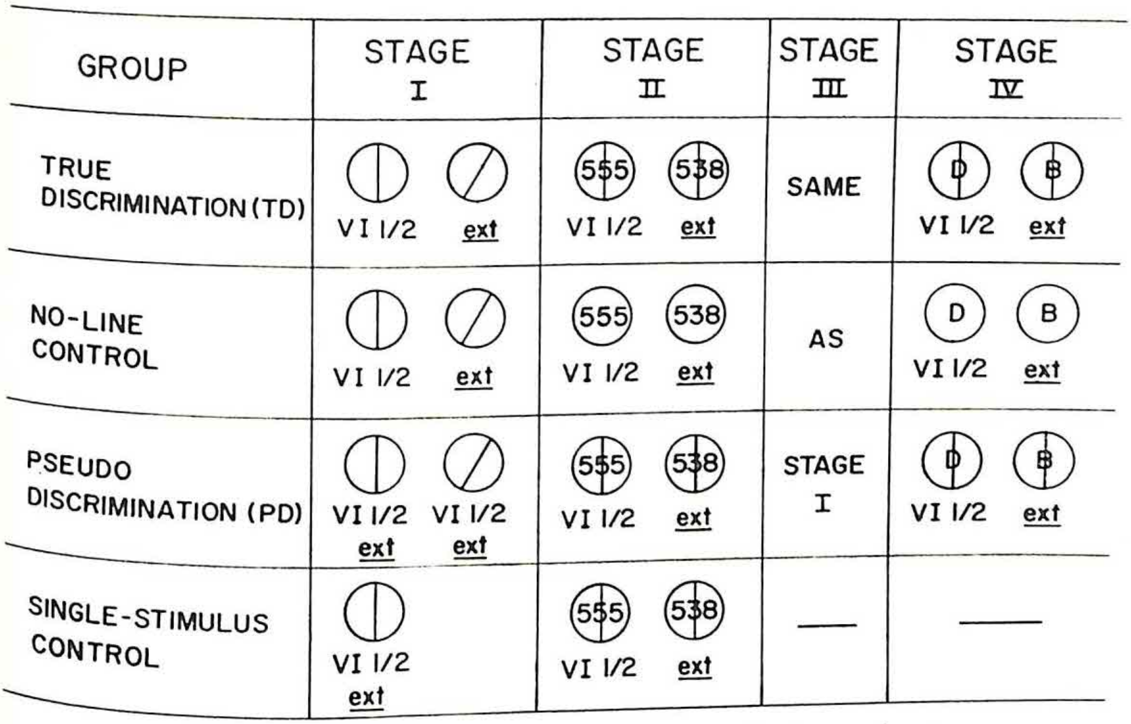


FIG. 1. Schematic representation of the conditions used in the experiment.

vertical line stimulus was superimposed on the two wavelengths; for the no-line control (NLC) group it was not. Since *S*s had been trained to attend to (i.e., respond on the basis of) line angle, it could be presumed that TD *S*s would continue to do so, to the detriment of their wavelength discrimination performance.

METHOD

Apparatus.—Two Grason-Stadler pigeon chambers were used. In-line display cells fitted with Kodak Wratten filters No. 74 and 99 provided wavelength stimuli of 538 and 555 mμ, respectively. Kodak neutral density filters (No. 96) with .10 and .40 density provided two different intensities of white light. The display cells also produced a white line ⅛ in. wide × ⅞ in. high on a black background, in either the vertical (90°) position or rotated 30° (clockwise) from vertical (i.e., at 60°). Miniature lamps (General Electric No. 44) were used in the display cells.

Procedure.—The procedure employed in the present study is schematically represented in Fig. 1. After attaining the 70–75% weight level, 10 birds were each randomly assigned to the TD, pseudo discrimination (PD), and NLC Groups. The single-stimulus control (SSC) group was run later and 12 birds were assigned to this group.

Each *S* was given magazine and key-peck training on the first day. Reward consisted of 3 sec. of access to a tray of mixed grain. On the next day, *S*s received 50 continuous reinforcements. Over the next 3 days, *S*s earned 30 reinforcements per day on a gradually lengthening temporal interval, such that by the end of the third day, a variable interval 1-min. (VI-1) schedule was in effect. During this preliminary training, the white vertical line on a black surround was continuously present on the response key.

Stage 1.—Stage 1 of training was begun on the sixth day. During this and subsequent stages, *S*s received daily training sessions of 30-min. duration in which stimulus-on periods of 50-sec. duration were separated by 10-sec. blackout periods. The TD and NLC groups were given discrimination training with randomly ordered presentations of the white vertical line (S+) and the 60° line (S−). During discrimination training there were three blocks of 10 stimulus presentations each per day, and within each block, S+ and S− were each presented five times. Responses were reinforced on a VI-½ schedule during S+ periods only. Discrimination training was continued until *S* had met the criterion ratio of 10 responses to S+ for each S− response for 3 days in a row.

Also during Stage 1, the PD group experienced the same sequence of the two training stimuli as did the TD group, but the stimuli were not correlated with the reinforcement contingency. For half the time (determined on a random basis), the VI-½ reinforcement schedule was in effect during presentations of the vertical line, and for the other half, reinforcement accompanied the 60° line. Each PD *S* was matched with a TD *S* and its training continued until that TD *S* had reached criterion.

The SSC group was presented with only the vertical white line during its stimulus-on periods, for half of which the VI-½ reinforcement schedule was in effect; for the other half (determined randomly), no reinforcement was available. Each *S* in this group was also matched with a TD *S*. One-half of the *S*s were given single-stimulus training until the TD *S* had reached criterion; the other half of the *S*s were given single-stimulus training for one-half the time required for the matched TD *S* to reach criterion.

Stage 2.—During Stage 2, the TD, PD, and SSC groups were given discrimination training with 555 mμ as S+ and 538 mμ as S−, with the vertical white line superimposed on both stimuli. Responses to S+ only were reinforced on VI-½ schedule. The NLC group was also given discrimination training with 555 mμ as S+ and 538 mμ as S−, but the vertical line was never superimposed on either stimulus for this group. In Stage 2 (and again in Stage 4), the criterion used was a single block of 10 trials during which there were more than 10 S+ responses to each S− response.

Stage 3.—Following completion of Stage 2, TD and NLC groups were returned to the Stage 1 discrimination between the white vertical line (S+) and the white line tilted at 60° (S−). Responses were again reinforced on a VI-½ schedule during S+ periods only. Discrimination training was continued until *S* had met the original criterion of 10 S+ responses to each S− response for 3 consecutive days. Following completion of the Stage 2 discrimination, PD *S*s were returned to nondifferential training between the vertical and tilted lines, each PD *S* being matched with a TD *S*. Training continued until the TD *S* had reached the criterion. Thus Stage 3 was a repetition of Stage 1 for three of the four groups. The SSC group was dropped after Stage 2.

Stage 4.—The TD, PD, and NLC *S*s were required to discriminate between two stimuli differing only in brightness. The dimmer value (S+) was produced by illuminating the light behind the .40 neutral density filter; the brighter value (S−) was produced by simultaneously illuminating the lights behind the .40 and the .10 filters. For the TD and PD groups, the vertical white line (previous S+) was superimposed on both stimuli and was clearly visible (to the human eye). For the NLC group, the white vertical line was never superimposed on either brightness value.

Subjects.—The *S*s were 42 experimentally naive domestic pigeons maintained at 70–75% of their ad libitum weights. During the course of the experiment, several *S*s became ill and were discarded. Nine TD, 9 PD, 12 SSC, and 10 NLC *S*s completed Stages 1 and 2. Nine TD, 6 PD, and 9 NLC *S*s completed Stages 3 and 4. Twelve *S*s were placed in the SSC condition (rather than 10) to provide a better comparison of the effects of two amounts of single-stimulus training in Stage 1 and to allow for possible *S* loss through illness. No such loss occurred, and thus 12 single-stimulus *S*s had to be matched to 9 TD *S*s. Five SSC *S*s received as much training as their (randomly) matched TD partners; 4 received half as many sessions as their partners. The remaining 3 SSC *S*s were matched to the mean of the TD group; 2 received half as many sessions, while the third received the mean number.

RESULTS

Stage 1.—In Stage 1, the TD and NLC groups gradually mastered the line angle discrimination task. The TD group met criterion in a mean of 11.5 sessions (including 3 criterion sessions); for the NLC group, 10.0 sessions were required. These two groups also performed approximately comparably over the entire course of the acquisition of the discrimination. The PD group, on the other hand, continued to respond equally to the two line angles. All four groups responded with approximately equal rate to the vertical line stimulus at the end of Stage 1 training.

Stage 2.—The data of greatest interest, of course, are those obtained from Stages 2 and 4. In Stage 2, for three of the four groups, no *S*s met criterion during the first nine blocks of 10 trials and so group data were plotted to this point. Two *S*s in the NLC group met criterion on Block 6. For these *S*s, the data points obtained on Block 6 were used as a conservative estimate of what performance on Blocks 7, 8, and 9 would have been. These estimated data points were included in the computations of the values plotted in Fig. 2a, 2b, and 2c, and in the associated analyses of variance.

Figure 2a presents the mean percentage of total responses emitted in the presence of the S+ stimulus by the four groups during the first nine blocks of 10 trials for the Stage 2 discrimination. Values plotted are group means calculated from individual percentages. As can be seen, the TD and NLC groups show a more rapid improvement in performance than do the PD and SSC groups. It can also be seen that the presence of the previous S+ stimulus during Stage 2 has virtually no effect on the acquisition of

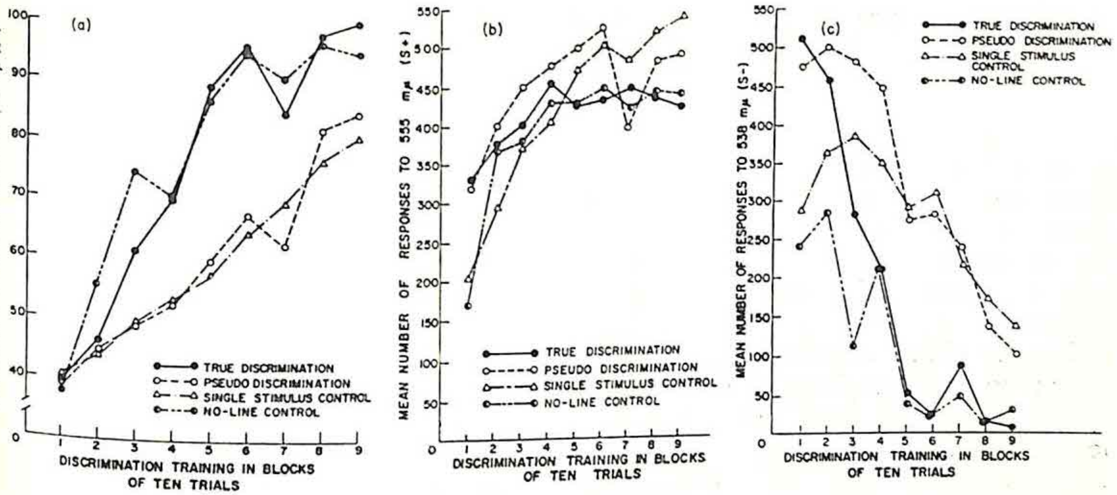


FIG. 2. Comparison of: (*a*) mean percentage of total responses to 555 mμ (S+), (*b*) mean number of responses to 555 mμ, and (*c*) mean number of responses to 538 mμ (S−) over nine blocks of 10 trials each during Stage 2 for the four groups of *S*s.

the discrimination, since the curve for the NLC group is essentially the same as that of the TD group. Similarly, the PD and SSC groups show essentially the same performance, indicating that experience with the two stimuli in Stage 1 for the PD group had little if any effect on their performance in Stage 2.

An analysis of variance performed on the percentage data yielded a significant between-groups difference, $F\ (3, 36) = 15.84$, $p < .01$. Since the groups all started at a comparable level, this indicates group differences in performance as a function of practice.

Another index of discrimination performance available was the number of trials required to reach the discrimination criterion. The TD and NLC groups required means of 57.77 and 55.00 trials, respectively. Corresponding means for the PD and SSC groups were 104.44 and 115.00. Because of heterogeneity of variance, a log transformation was used, with analysis of variance then yielding $F\ (3, 36) = 15.52$, $p < .01$.

It should be remembered that *S*s in the SSC group were matched to those in the TD group, and six of these *S*s were trained for as many sessions as the paired TD *S* required to reach the Stage 1 criterion, and the other six received half as much training. This was done because there were two equally plausible ways to match the SSC *S*s with the others in Stage 1: (*a*) in terms of total training time (number of sessions), and (*b*) in terms of training time with the vertical line stimulus. The use of both procedures (with half of the *S*s in each) permitted their subsequent comparison. The most obvious comparison is simply between the half of the *S*s which received the most Stage 1 training and the half which received the least. As could be expected, the former subgroup responded somewhat (although nonsignificantly) more to both wavelength stimuli, particularly during the first few sessions of Stage 2 training, but the two subgroups were entirely comparable in percentage of total responses to S+ and in mean trials to criterion. Thus absolute level of responding in Stage 1 and in Stage 2 does not appear to be related to the rapidity of the acquisition of the wavelength discrimination.

The differences in performance illustrated in Fig. 2a could reflect a more rapid increase in response rate to S+ for the TD and NLC groups, a more rapid decrease in response rate to the S− stimulus for these two groups, or both. To select among these alternatives, response rates to the two training stimuli were analyzed separately over the nine blocks of trials.

Figure 2b presents the mean number of responses to the S+ stimulus for the four groups over the nine blocks of trials during Stage 2. The SSC and NLC groups had a considerably lower rate of response to S+ on the first block of trials than did the other two groups, but beyond this point group differences were negligible. In the NLC group, the lowered rate may simply be generalization decrement as a consequence of the removal of the line, which was present in prior training. Half of the *S*s in the SSC group had received only half as much VI training as did their matched partners in the TD group, which explains why the mean response rate in this group would initially be lower than the others. There is, unfortunately, no obvious explanation of the precipitous dip in response rate to S+ in the PD group in Block 7.

An analysis of variance performed on these data yielded a significant trials effect, $F\ (8, 288) = 20.86$, $p < .01$, and a significant Group × Trials interaction, $F\ (24, 288 = 2.64$, $p < .01$, but no significant overall group differences ($F < 1$).

Figure 2c presents the mean number of responses to the S− stimulus over the nine blocks of trials for the four groups during Stage 2. An analysis of variance performed on these data showed a significant group effect, $F\ (3, 36) = 13.24$, $p < .01$, as well as a significant trials effect, $F\ (8, 288) = 56.81$, $p < .01$, and a Group × Trials interaction, $F\ (24, 288) = 5.07$, $p < .01$. The preceding analyses indicate that the performance differences exhibited among the groups in Fig. 2a are primarily a function of a more rapid extinction of responses to the S− stimulus on the part of the TD and NLC groups.

Stage 3.—In Stage 3 of training, the Stage 1 training conditions of the TD, PD, and NLC groups were reinstated. The *S*s quickly resumed their prior levels of performance under these conditions. On the first day of Stage 3 training, the TD and NLC groups emitted approximately 90% of total responses to the vertical line S+ whereas the PD *S*s responded equally to both line angles. The TD group required a mean of 4.88 sessions to reachieve criterion, compared to 3.55 for the NLC group. Response levels to the two line angles were also comparable to what they had been at the end of Stage 1.

Stage 4.—To present a group picture of performance in Stage 4, it was possible to include data from only six blocks of 10 training trials each. Two *S*s in the TD group met criterion in Blocks 2 and 3, respectively, and data for the remaining blocks were estimated for these *S*s.

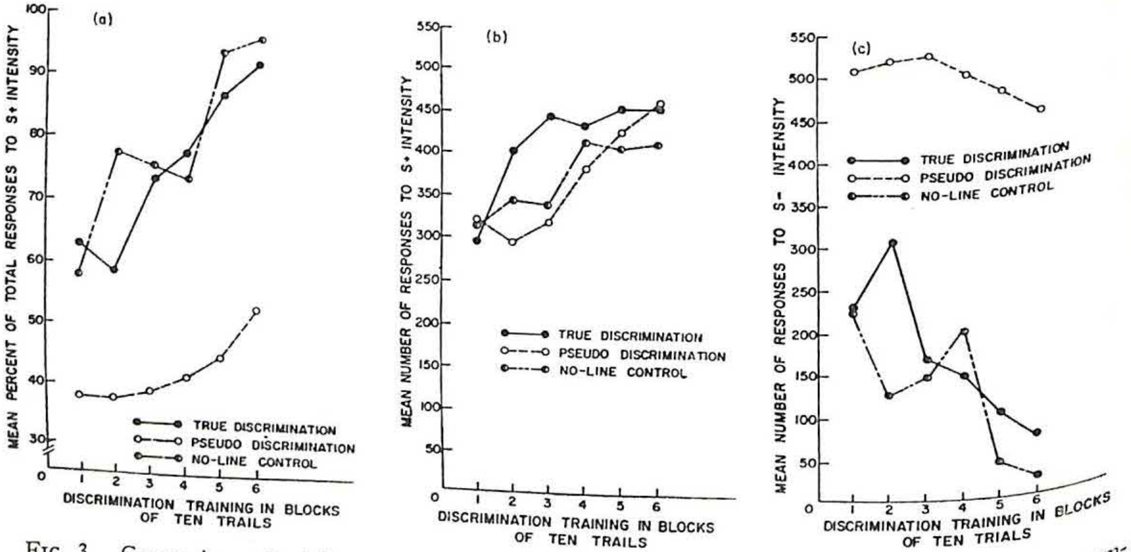


FIG. 3. Comparison of: (*a*) mean percentage of total responses to the S+ intensity, (*b*) mean number of responses to the S+ intensity, and (*c*) mean number of responses to the S− intensity over six blocks of 10 trials each during Stage 4 for true discrimination, pseudo discrimination, and no-line control groups.

Figure 3a presents the mean percentage of total responses emitted in the presence of the S+ stimulus by the three groups during the first six blocks of 10 trials for the Stage 4 discrimination. Values plotted are group means calculated from individual percentages. Again the TD and NLC groups show a more rapid improvement in performance than does the PD group, and the difference would seem to be greater than in Stage 2. In fact, the TD and NLC groups learned the brightness discrimination so rapidly that approximately 60% of total responding was made to S+ on the very first block of 10 trials. The performance of the PD group in Stage 4 was extremely poor, hovering around 40% of total responses to S+ for several sessions. This reflects both the time spent in feeding and pauses of a few seconds which typically followed feeding. Of course, neither of these occurred to interrupt key pecking during S− periods. As was true in Stage 2, the presence of the previous S+ stimulus has little or no effect on the performance of *S*s since the NLC and TD groups show approximately the same level of performance.

An analysis of variance performed on the percentage data yielded a significant group difference, $F\ (2,\ 21) = 25.61$, $p < .01$, a significant trials effect, $F\ (5,\ 105) = 23.26$, $p < .01$, and a significant Group × Trials interaction, $F\ (10,\ 105) = 2.65$, $p < .01$.

The TD and NLC groups required means of 47.77 and 42.22 trials, respectively, to achieve the 10/1 S+/S− discrimination criterion. The PD group needed a mean of 330.00. Because of heterogeneity of variance, a log transformation was used, with a subsequent analysis of variance revealing $F\ (2,\ 21) = 31.78$, $p < .01$.

Figure 3b presents the mean number of responses to the S+ stimulus for the three groups over the first six blocks of 10 trials during the Stage 4 discrimination. An analysis of variance performed on these data showed a significant trials effect, $F\ (5,\ 105) = 10.09$, $p < .01$, but no significant group differences or Group × Trials interaction. Thus, the differences exhibited in the percentage data are not a function of an increased rate of response to the S+ stimulus by the TD and NLC groups.

Figure 3c presents the mean number of responses to the S− stimulus for the three groups over the first six blocks of 10 trials during the Stage 4 discrimination. An analysis of variance performed on these data showed a significant between groups difference, $F\ (2,\ 21) = 22.10$, $p < .01$, and a significant trials effect, $F\ (5,\ 105) = 6.48$, $p < .01$, but no Group × Trials interaction, $F\ (10,\ 105) = 1.65$. There was considerable within-group variability in the data plotted in Fig. 3b and 3c, such that in neither case was there a significant Group × Trials interaction. What is most important to note about Fig. 3b is that by the end of training, the three groups did not differ in level of responding to the S+. As shown in Fig. 3c, however, the groups differed in level of responding to the S−, even in the first block of 10 trials. On this basis it may be concluded that again (as in Stage 2) the performance differences between the TD and NLC groups, on the one hand, and the PD group, on the other, are a function of the less rapid extinction of responses to S− in the latter group.

DISCUSSION

These results are entirely consistent with the proposition that in comparison with single-stimulus and nondifferential training, extradimensional discrimination training produces a state of general attentiveness, such that the formation of a subsequent discrimination orthogonal to the first is facilitated. It should be emphasized that the line angle training dimension used in Stages 1 and 3 was orthogonal to the dimensions of wavelength and brightness used in Stages 2 and 4. The line angles were equivalent in wavelength (both were white) and in brightness, with the possible exception of of unsystematic differences due to aging and replacement of bulbs in the different chambers.

The superficial similarity of these findings to those ascribed to the formation of discrimination learning sets by Harlow (1949) is so obvious that it seems appropriate to point out some basic differences. Unlike a typical learning set experiment, in the present study successive discrimination training was used rather than simultaneous, and substantial training was given on a single problem, rather than a little training on a great many. The magnitude of the obtained facilitative effect seems comparable to that produced in primates by experience with a very great number of prior problems.

The data from Stage 2 suggest that the observed effect was due entirely to an enhancement of the performance of the discrimination-trained groups; the PD group behaved no differently from the SSC group. This finding is in agreement with that of one of the experiments reported by Thomas et al. (1969) in which the generalization gradients of *S*s given extradimensional TD training were reliably steeper than those of PD and SSC *S*s which, in turn, did not differ. The results obtained in Stage 4, however, suggest that PD training can have an inhibiting effect. Although there was no SSC group with which a direct comparison can be made, it should be noted that while the TD and NLC groups required fewer trials to reach criterion on the Stage 4 (brightness) problem than they had on the Stage 2 (wavelength) discrimination, the PD group required more than three times as many. Further research would be needed to determine whether an inhibiting effect of PD training cumulates after two different PD experiences, occurs with some stimulus dimensions but not with others, or is unreliable altogether. On the other hand, the results of Stages 2 and 4 agree in suggesting that the major benefit of prior discrimination training is a more rapid extinction of responses to the subsequent S−.

Finally, it should be noted that there was no significant difference in the performance of the TD and NLC groups in either Stage 2 or Stage 4 of the present study. Yet there is no reason to doubt that having been trained to respond on the basis of line angle, the TD *S*s would continue to attend to this stimulus (at least at the outset) in Stages 2 and 4. It must be concluded, therefore, that attention paid to line angle does not detract from attention paid to wavelength or brightness. This finding contradicts a basic assumption of selective attention theory that a negative relationship *must* attain between attention paid to one stimulus aspect and attention paid to others. Although there is ample evidence to suggest that such a negative relationship can exist (cf., e.g., Mackintosh, 1965; Trabasso & Bower, 1968), it is becoming increasingly clear that this is not invariably the case.

REFERENCES

Harlow, H. F. The formation of learning sets. *Psychological Review,* 1949, **56**, 51–65.

Honig, W. K. Attentional factors governing the slope of the generalization gradient. R. M. Gilbert & N. S. Sutherland (Eds.), *Animal discrimination learning.* London: Academic Press, 1969. (a)

Honig, W. K. Attention and the modulation of stimulus control. In D. I. Mostofsky (Ed.), *Attention: Contemporary studies and analyses.* New York: Appleton-Century-Crofts, 1969, in press. (b)

Lawrence, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology,* 1950, **40**, 175–188.

Mackintosh, N. J. Selective attention in animal discrimination learning. *Psychological Bulletin,* 1965, **64**, 124–150.

Reinhold, D. B., & Perkins, C. C., Jr. Stimulus generalization following different methods of training. *Journal of Experimental Psychology,* 1955, **49**, 423–427.

Sutherland, N. S. The learning of discriminations by animals. *Endeavor,* 1964, 23, 140–152.

Thomas, D. R. The use of operant conditioning techniques to investigate perceptual processes in animals. In R. M. Gilbert & N. S. Sutherland (Eds.), *Animal discrimination learning.* London: Academic Press, 1969.

Thomas, D. R., Freeman, F., Svinicki, J. G., Burr, D. E. S., & Lyons, J. The effects of extra-dimensional training on stimulus generalization. *Journal of Experimental Psychology,* 1969, in press.

Trabasso, T., & Bower, G. H. *Attention in learning.* New York: Wiley, 1968.

(Received February 8, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 163–167

VERTICAL EYE MOVEMENT AND SPACE PERCEPTION: A DEVELOPMENTAL STUDY

DONALD H. THOR,[1] JOHN J. WINTERS, JR., AND DAVID L. HOATS[2]

E. R. Johnstone Training and Research Center, Bordentown, New Jersey

Children varying in age and intelligence viewed electroluminescent lamps in an arc of 10–90° without head movement in a dark room. When identical and equidistant lamps were presented in pairs, the higher lamps were judged smaller and more distant than the lower lamps. Visual space is apparently altered with eye movement in the vertical plane.

Although nearly equidistant from *O* in both horizontal and vertical viewing planes, the sun or moon is generally perceived as larger and nearer when viewed on the horizon. Kaufman and Rock (1962a, 1962b; Rock & Kaufman, 1962) have claimed experimental support for an explanation proposed by Ptolemy in the second century A.D.; the illusion is said to be primarily dependent on the absence or presence of terrain. Although generally accepted and considered the most consensual explanation by a number of historically eminent scientists (Helmholtz, 1925; Reiman, 1902; Smith, 1738), the issue has not been entirely resolved for a variety of cogent reasons discussed by Wallach and Boring (1962).

Since the illusory effect has been obtained in a variety of experimental situations in which terrain cues were excluded or minimal (Bilderback, Taylor, & Thor, 1964; Kammann, 1967; Leibowitz & Hartman, 1959; Schur, 1925; Thor & Wood, 1966), it may tentatively be concluded that the illusory effect is a general characteristic of human space perception and not restricted to the isolated case of viewing the moon or sun.

The present experiment was conducted to test the relative visual distortion in four developmental populations when equidistant visual stimuli were viewed at varying elevations, permitting eye movement and restricting any head movement.

METHOD

Subjects.—Three groups of normal children of varying ages and a group of retarded adolescents were tested. See Table 1 for group characteristics.

Mental age (MA) and chronological age (CA) were both considered in view of a report by Leibowitz and Hartman (1959) describing a greater illusory effect in children than in adults. A retarded group was also included in order to determine whether level of intelligence affected the illusion. Boys and girls were equally represented in all groups.

Apparatus and procedure.—All *S*s with the exception of the kindergarten group were screened for binocular, far visual acuity with a Bausch and Lomb ortho-rater. Only *S*s with Snellen equivalents of 20/40 (uncorrected) or better were included. One kindergarten *S* and two retarded *S*s normally wore glasses; due to possible distortion of stimuli through lenses or frames of the eyeglasses, they were not worn during the experiment. All *S*s were also screened for understanding of the following concepts: larger, smaller, nearer, farther, top, and bottom. Two sizes of rubber balls in varying positions were used. Each *S* was required to state the appropriate relationships when questioned. Three retardates and 27 kindergarten children were dropped from the experiment and replaced due to failure to understand the requisite concepts.

After screening for visual acuity and conceptual ability, each *S* was led blindfolded into a lightproof testing room, seated in a lounge chair with back tilted at a 45° angle, and carefully positioned so that the eyes were equidistant from visual targets located on a curved Masonite surface in a 90° arc from horizontal to zenith. The surface was 61 cm. wide and 213 cm. from the eyes of *O*. When the apparatus was placed on a table, the counterweighted surface extended over *S* as a cantilever.

---

[1] Requests for reprints should be sent to Donald H. Thor, Edward R. Johnstone Training and Research Center, Bordentown, New Jersey 08505.

[2] For assistance in obtaining Ss, the authors wish to express appreciation to G. Reuben, R. Schmidt, E. Banos, and W. Gretsch of the Willingboro, New Jersey, schools. The authors also thank W. Henry, R. Bice, F. Woods, and M. Dougherty for their help with the younger children and the research staff at Johnstone for their critical reading of the manuscript.

TABLE 1

Characteristics of Groups

Group	N	Mean IQ	SD	Mean CA	SD
Kindergarten	32	—	—	5.59	.78
Equal MA	32	109.25	8.87	11.08	.38
Equal CA	32	113.38	11.09	16.91	.32
Retardates	32	65.78	7.95	15.59	1.66

Note.—IQ scores for the kindergarten group were not available.

Targets were circular electroluminescent panels (Sylvania) subtending 1°32′ and 0°41′ of visual arc (58- and 24-mm. diameters). Operated at 50 v. ac and 60 Hz., these lamps produced a uniform, diffuse blue light with a peak spectral luminosity at 410 mμ and brightness of .2 mL. measured at the surface. The centers of the larger targets were located at 10°, 30°, 50°, 70°, and 90° from the horizontal; smaller targets were centered at 20°, 40°, 60°, and 80° elevation.

After removal of the blindfold, each *S* was presented with the 10° and 90° targets and asked to position his head so that he could see both targets by moving only his eyes. The primary position was therefore approximately 50° elevation, which allows distinct vision of targets at 10° and 90° according to Taylor and Boring (1942). It was necessary for *S* to move his eyes to compare any two targets. An elastic band was then placed over his forehead, and *E* placed a hand on the headband to detect any subsequent head movement. A second *E* in an adjoining room operated a switching console presenting paired targets for 5-sec. exposure and recording verbal responses of *S*. The task was to look at two illuminated targets and report the larger, smaller, farther, or nearer target by indicating "top" or "bottom." Half of the *S*s in each group were given one size followed by one distance instruction; the other half received the opposite order.

There were thus eight combinations of instructions and order: larger–nearer, larger–farther, smaller–nearer, smaller–farther, nearer–larger, farther–larger, nearer–smaller, and farther–smaller. Larger and smaller response instructions were counterbalanced, as were response instructions to nearer and farther. Order of size and distance judgment was also counterbalanced, with two orders of 10 test pairs presented for each response condition. Following five training trials to familiarize *S* with the task, *S* responded to the experimental trials under size or distance instructions. He then received the alternate instructions, five training trials, and the experimental trials. Each *S* made a total of 48 responses, 24 to size and 24 to distance (4 lie items and 20 test items in each condition). In cases of indecision, *S* was requested to "guess." Intertrial interval was approximately 5 sec.

Presentation of two randomized orders of all possible pairs of the five larger discs were made with interspersed lie items. A lie item was one of the small targets paired with one of the larger targets. Failure on lie items resulted in dropping and replacing one equal-CA, two equal-MA, six retarded, and five kindergarten *S*s.

Results

All test trial responses in the illusory direction, i.e., lower target nearer and larger or upper target farther and smaller, were tabulated. No significant differences were found among order, size instruction (larger–smaller), and distance instruction (nearer–farther). Therefore, the data were collapsed across these variables.

Reliability measures within groups were computed for size and distance judgments. Since each of 10 possible comparisons of the five test stimuli (large discs) were repeated, a tabulation was made of the number of agreements for each comparison. Single-mean t tests were then computed to determine if agreement was beyond chance significance within groups and instruction. All eight t values (4 groups × 2 instructions) were highly significant ($p < .001$), indicating that within groups and instructions, *S*s were consistent in their judgments.

Single-mean t tests for illusory effect by sex, instruction, and group were computed (collapsed across the 10 stimulus comparisons): Of the 16 tests of significance, 7 were significant at $p < .001$, 2 at $p < .01$, and 1 at $p < .05$. Of the remaining, 3 were marginally significant ($p < .10$). Nonsignificant ($p > .10$) illusory effects were obtained in retarded and kindergarten female groups on size judgments and equal-MA females on distance judgments. When the data were collapsed further across instruction, 6 t tests were highly significant ($p < .001$), 1 was significant ($p < .05$), and 1 (kindergarten females) was not significant ($p < .30$). The preceding analyses indicate that the illusory effect was obtained in all groups, by males and females (excepting kindergarten females), and with both size and distance instructions.

A 4 × 2 × 2 analysis of variance was performed on the individual scores. Groups.

sex, and instructions variables were not significant. The Groups × Sex interaction was significant, $F\ (3,\ 120) = 5.42,\ p < .01$.

Two separate one-way analyses of variance were computed for males and for females by groups. Males did not differ significantly, whereas the females did, $F\ (3,\ 60) = 6.09,\ p < .005$. A Duncan multiple-range test showed that the equal-CA females demonstrated significantly more ($p < .05$) illusory effect than equal-MA females, who did not differ from the retarded females and all of whom differed significantly ($p < .05$) from the kindergarten females.

Figure 1 displays the relative extent of illusory effect by groups and sex. The significant Sex × Groups interaction can be seen as primarily due to the lack of illusion obtained by kindergarten females.

Further analyses were made to investigate the importance of target elevation. A one-way analysis of variance was computed for target separations of 80°, 60°, 40°, and 20°, collapsing data across sex, groups, and instructions. A significant difference was found, $F\ (3,\ 508) = 10.67,\ p < .001$, and a Duncan multiple-range test revealed that $80° > 60° > 40°$ ($p < .05$) and that the 40° and 20° comparison did not differ in contributing to magnitude of illusion.

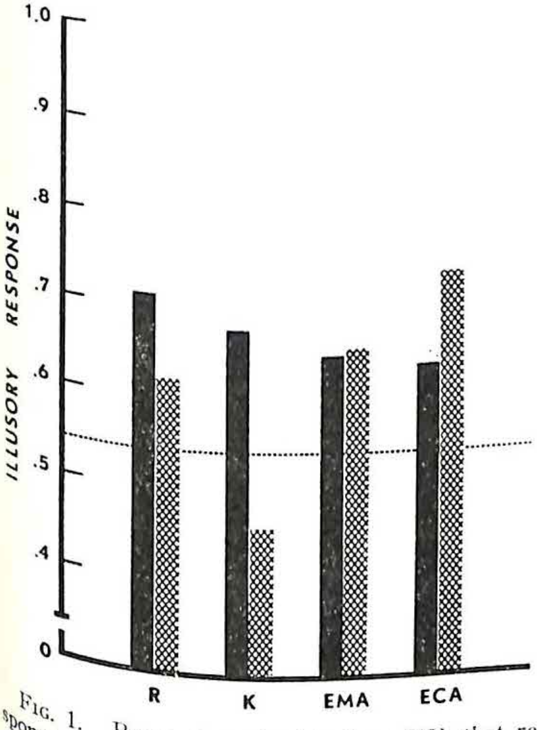


FIG. 1. Proportion of trials ($n = 640$) that response was in the illusory direction by group (retardate, kindergarten, equal-MA, equal-CA) and sex (males—solid bar, females—hatched bar). (Horizontal dashed line indicates significant positive deviation, $p < .05$, from theoretical mean when $p = \frac{1}{2}$ and $q = \frac{1}{2}$, ordinate value of .539.)

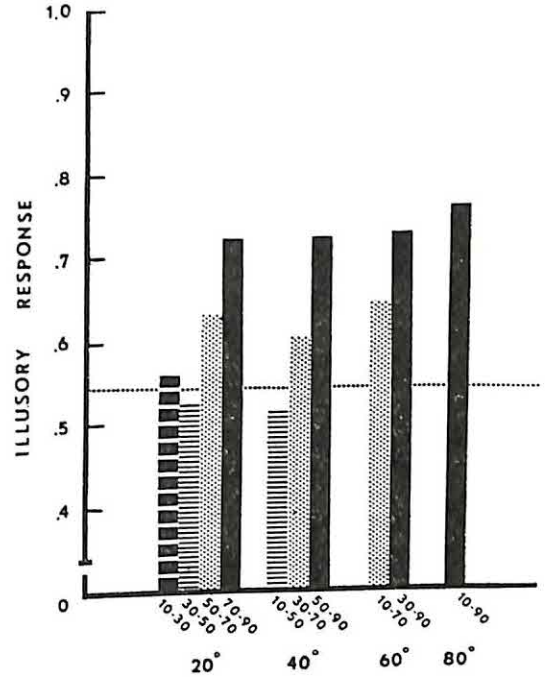


FIG. 2. Proportion of trials ($n = 512$) that response was in the illusory direction by angle of separation for each of the 10 possible pairings of the five test stimuli. (Data collapsed over groups, sex, and viewing instruction. Significant, $p < .05$, positive deviation from chance mean indicated by horizontal dashed line, ordinate value of .543.)

Three two-way analyses of variance (Groups × Elevation) were computed for 60°, 40°, and 20° separations to examine significance of elevation of paired targets. The two 60° separations (10–70° and 30–90°) differed significantly, $F\ (1,\ 124) = 10.78,\ p < .001$; the three 40° separations (10–50°, 30–70°, and 50–90°) differed significantly, $F\ (2,\ 248) = 30.20,\ p < .001$; and the four 20° separations (10–30°, 30–50°, 50–70°, and 70–90°) differed significantly, $F\ (3,\ 372) = 18.95,\ p < .001$. None of the within-group comparisons nor any of the interactions were significant.

The significant elevation comparisons are described in Fig. 2. With the exception of the 30–50° comparison, it can be seen that the illusory effect is greater when the target comparisons are closer to the zenith, regardless of the extent of separation.

No significant differences among groups negate intelligence, chronological age, and mental age as significant variables in the present study. An overall sex difference and instructions for responding to size and distance may also be considered nonsignificant variables in the present investigation.

DISCUSSION

The present design does not exclude experiential variables as contributing to the distortion effect since previous viewing experience and expectation of ceiling height were not controlled. Epstein (1967) has claimed that suggestions of distances involved can influence visual perception in reduced cue situations. Since kindergarten boys do experience the illusion comparable to older boys, experiential variables would appear to have slight influence after 5 yr. of age. In the present investigation, the discrepant behavior of kindergarten girls may be attributed to fear of the dark room or an apprehensive reaction to strangers in an unusual situation.

The results appear inconsistent with those of Leibowitz and Hartman (1959), who reported a greater reduction of the zenith object by children. They interpreted their results in terms of the development of size constancy, i.e., the children's rate of development of size constancy is slower in the vertical than in the horizontal direction. Their results, however, are not necessarily in conflict with the present results. Leibowitz and Hartman required their *S*s to make size judgments of distant objects, outdoors and indoors. It is possible that in their indoor condition the cues to depth were sufficient to permit size constancy to occur. That condition resulted in less of a difference between the youngest and oldest *S* than their outdoor condition. The present cue-deprived environment revealed no differences. Therefore, it is possible that size constancy is influential in a full-cue situation, but diminishes as cues are reduced until no cues remain, whereupon eye movement is sufficient to cause the illusory effect. This explanation would also be consistent with Holway and Boring's (1941) explanation of the reduction of size constancy.

With reference to previous work relating size constancy to chronological age, several investigators (Beyrl, 1926; Frank, 1926; Zeigler & Leibowitz, 1957) have claimed that in full-cue situations, size constancy develops until approximately 10 yr. of age, when it is equal to that of adults. Considerable individual variability exists, and the results may be somewhat confounded with task instruction. More recently, Bower (1965) has found three-dimensional size constancy in infants 40–60 days old. A more definitive study than the present would require a comparison of reduction and nonreduction of cues, a range of viewing distances including distances 10 or 20 times greater, and a greater range of mental and chronological ages.

The present work in a laboratory situation serves to further validate the claim that the so-called moon illusion effect can be produced without the moon, terrain, or multitude of associated cues present when experiments are conducted outdoors. The addition of depth cues apparently intensifies the effect.

According to Rock and Kaufman (1962), "Eye-elevation does not account for the moon illusion (nor, for that matter, does head elevation) [p. 1031]." They have also reported that a terrain is critical for the illusory effect and that only a minute illusory effect may be demonstrated in a completely dark field. Although the present data fail to support their position on the moon illusion, they do support an earlier position taken by Boring (1943) and his students: "this phenomenon depends upon the physiological properties of the observer and not upon physical factors external to him . . . [p. 56]." And also, "It depends on raising or lowering the eyes [p. 59]."

Of major importance is the fact that human visual space perception appears to vary with eye movement in the vertical plane. Size and distance judgments reveal that the perceived object looks smaller and more distant when elevated. The discrepancy between stimuli of identical size and at identical distances from the eyes of *O* is more apparent when viewed in the near vertical.

REFERENCES

BEYRL, F. Über die Grössenauffassung bei Kindern. *Zeitschrift für Psychologie*, 1926, **100**, 344–371.

BILDERBACK, L. G., TAYLOR, R. E., & THOR, D. H. Distance perception in darkness. *Science*, 1964, **145**, 294–295.

BORING, E. G. The moon illusion. *American Journal of Physics*, 1943, **11**, 55–60.

BOWER, T. G. R. Stimulus variables determining space perception in infants. *Science*, 1965, **149**, 88–89.

EPSTEIN, W. Perceived distance in imagined space. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 341–343.

FRANK, H. Untersuchung über Sehgrössenkonstanz bei Kindern. *Psychologische Forschung*, 1926, **70**, 137–145.

HELMHOLTZ, H. V. *Physiological optics.* (J. P. C. Southall, Ed.) Vol. 3. New York: Optical Society of America, 1925.

HOLWAY, A. H., & BORING, E. G. Determinants of apparent visual size with distance variant. *American Journal of Psychology*, 1941, **54**, 21–37.

KAMMANN, R. The overestimation of vertical distance and slope and its role in the moon illusion. *Perception and Psychophysics*, 1967, **2**, 585–589.

KAUFMAN, L., & ROCK, I. The moon illusion, I. *Science*, 1962, **136**, 953–961. (a)

KAUFMAN, L., & ROCK, I. The moon illusion. *Scientific American*, 1962, **207**, 120–130. (b)

LEIBOWITZ, H., & HARTMAN, T. Magnitude of the moon illusion as a function of the age of the observer. *Science*, 1959, **130**, 569–570.

REIMAN, E. Die scheinbare Vergrösserung der Sonne und des Mondes am Horizont. *Zeitschrift für Psychologie*, 1902, **30**, 1–38, 161–195.

ROCK, I., & KAUFMAN, L. The moon illusion, II. *Science*, 1962, **136**, 1023–1031.

SCHUR, E. Mondtäuschung und Sehgrössenkonstanz. *Psychologische Forschung*, 1925, **7**, 44–80.

SMITH, R. *A compleat system of opticks.* Vol. 1. Cambridge, England, 1738.

TAYLOR, D. W., & BORING, E. G. The moon illusion as a function of binocular regard. *American Journal of Psychology*, 1942, **55**, 189–201.

THOR, D. H., & WOOD, R. J. A vestibular hypothesis for the moon illusion. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1966.

WALLACH, H., & BORING, E. G. On the moon illusion. *Science*, 1962, **137**, 900–901.

ZEIGLER, H. P., & LEIBOWITZ, H. Apparent visual size as a function of distance for children and adults. *American Journal of Psychology*, 1957, **70**, 106-109.

(Received February 10, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 168–174

POSTSTIMULUS OUTPUT SPECIFICATION AND DIFFERENTIAL RETRIEVAL FROM SHORT-TERM MEMORY

WILLIAM EPSTEIN[1]

University of Wisconsin

On each of eight trials, different pairs of lists of unrelated words and numbers were presented once with instructions to memorize both lists. Recall instructions given *after* the last item of the second input specified one of four recall outputs: O—numbers only or words only; F—numbers first, then words; S—words first, then numbers. Experiments I–III showed that for the second input, recall is superior under Cond. O. Experiments IV and V showed that the superiority of Cond. O is not present if the words and numbers are input cotemporally in a mixed list. The findings are interpreted as evidence that independently of acquisition variables, there are events during retrieval that affect short-term recall significantly.

The experiments reported in the present paper were suggested by an incidental finding in an earlier study (Epstein, 1969, Exp. II). In that study, *S*s were required to memorize an initial five-word string and an immediately succeeding list of eight unrelated words. For different *S*s the initial strings were either ordinary sentences, anomalous strings (grammatical but semantically anomalous), or ungrammatical random strings. Each of eight completely different string-list combinations was presented aurally only once, and immediately following the last word on the list, *S* was tested for free recall. The critical features of the procedure were the temporal position and specification of the recall instructions. The prescribed recall output varied from trial to trial and was specified only *after* the learning presentation was completed, i.e., immediately after the last word in the list. On each of two of the eight trials the following outputs were solicited: (*a*) "list only"; (*b*) "list, then string"; (*c*) "string, then list"; (*d*) "string only."

For all three types of preceding string, the number of list words recalled was dependent on the recall output specification. The "only" (O) condition yielded superior recall. Recalling the list second (S) yielded poorest recall, and recalling the list first (F) yielded recall that was intermediate between O and S.

Lower recall under Cond. S may be attributed to the proactive interference from prior recall activity, or to trace decay. What was unexpected was the consistent and significant superiority of list recall under Cond. O relative to Cond. F. This difference in recall cannot be attributed to differences in original learning. Attention and encoding may be presumed to have been the same on all learning trials since *S* could not know in advance which output would be specified. Nor can the difference be attributed to overt recall activity, or trace decay resulting from delay of recall. The sole difference that can be singled out is that under Cond. F, *S* had to keep the string in mind while recalling the list, but under Cond. O, *S* could disregard the string. More is said about this in the Discussion section.

EXPERIMENT I

The purpose of Exp. I was to determine if the original findings would be obtained again in a new experiment with number lists replacing the strings. In addition, Exp. I varied input order as well as output order. The main objectives of the original experiment required that the input always be in the order: string followed by list. In the present experiment, one group of *S*s had input order: words, then numbers (W_1N_2); and a second group had input order: numbers, then words (N_1W_2). For each group, on an equal number of trials, one of four recall outputs was specified after the learning trial: (*a*) "words only"; (*b*) "numbers only"; (*c*) "words, then numbers"; (*d*) "numbers, then words." The comparisons of interest were between recall output conditions, input conditions, and the Output × Input interaction.

---

[1] The author is indebted to Frank Shontz and S. M. Ebenholtz for useful discussions of this work.

Requests for reprints should be sent to William Epstein, Department of Psychology, University of Wisconsin, Psychology Building, Charter at Johnson, Madison, Wisconsin 53706.

Method

Materials.—The materials were eight lists of eight two-syllable words and eight lists of eight two-digit numbers. The 64 words were selected from Howes' (1966) tabulation of the frequency of occurrence of words in the spoken language of university students. They ranged in frequency from 5 to 8 per 100,000. All eight lists had the same mean frequency. The 64 numbers were selected from Battig and Spera's (1962) tabulation of the association value (AV) of numbers. They ranged in AV from .72 to 2.03. The mean AVs for the eight lists ranged from 1.33 to 1.36. Three different pairings of word lists and number lists were prepared.

Subjects.—The *S*s were 100 students who elected to participate in the experiment to satisfy a course requirement.

Procedure.—The *S*s were tested in groups of five. Ten groups ($N = 50$) were assigned to each of two input orders: W_1N_2—word list followed by number list; N_1W_2—number list followed by word list. Four groups in each input condition learned one of the sets of word-list–number-list pairings; the remaining six groups were evenly divided between the other two arrangements.

The instructions and learning lists were tape-recorded and presented aurally. The instructions informed *S* that there would be a series of learning trials, each presenting totally different lists, and that each learning trial would be followed by a recall test. On each trial, his task was to learn as much as possible of the two lists. The order of input, e.g., list of words followed by a list of numbers, was specified in the instructions. The *S* was informed that the recall test would require one of four outputs and that the outputs would be specified after the last item on the second list. The alternative output prescriptions were explained in detail and the cue words soliciting the various outputs were available for *S*'s inspection on a blackboard during the experiment. The outputs were specified on the tape by the statements: "words only"; "numbers only"; "words, then numbers"; "numbers, then words." Each of the output instructions was used on two of the eight trials. This latter information was not supplied in the instructions, nor was *S* told how many trials to anticipate. Within any particular list, recall order was unconstrained; the instructions informed *S* that he was free to recall the items from the designated list in any order he wished.

The items on the list were presented on tape at the rate of one each 1.5 sec. There was a 6-sec. pause between lists. The recall cue followed the last item on the second list after an interval of 4.5 sec. One minute was allowed for each recall test. On the recall test, *S* wrote his responses on a prepared sheet which contained two columns of eight underlined spaces, one headed "Words" and the other "Numbers." Sixteen test sheets were provided, although only eight were required. Each recall sheet was removed from view at the completion of the recall test. A new learning trial began 10 sec. after the conclusion of the recall test.

TABLE 1

MEAN NUMBER OF WORDS (WS) AND NUMBERS (NS) RECALLED FOR TWO INPUT ORDERS AND THREE RECALL OUTPUT SPECIFICATIONS: EXP. I

Input	Only		First		Second	
	M	*SD*	*M*	*SD*	*M*	*SD*
	Word output					
W_1N_2	4.05	1.03	4.18	1.19	4.00	1.26
N_1W_2	5.64	.73	5.00	.91	3.93	1.09
	Number output					
N_1W_2	2.97	1.11	2.32	1.05	2.43	1.10
W_1N_2	3.62	.98	2.84	1.03	2.45	1.04

Note.—Total possible = 8.

Results

Table 1 shows the mean number of items recalled for words and numbers separately. Examination of Table 1 shows the following: (*a*) Words were recalled better than numbers. (*b*) Under Cond. O and F, recall was higher from the list presented second than from the list presented first; but under Cond. S, recall was the same for items from the lists presented first and second. This was true both for words and numbers. (*c*) When the recalled items were from the list put in second, the amount recalled varied with the recall output specification as in the original experiment: O > F > S. This was true for both words and numbers. (*d*) When the recalled items were from the list put in first, amount recalled did not vary systematically with the recall output specification. This was most clearly the case for words.

Separate analyses of variance were computed for word recall and number recall following Winer's (1962, pp. 302–307) model for the two-factor experiment with repeated measures on one factor.

Words.—The main effects of input and output were significant: input, F (1, 98) = 27.68, $p < .001$; output, F (2, 196) = 25.81, $p < .001$. The Input × Output interaction was also significant, F (2, 196) = 20.17, $p < .001$. A series of individual F tests (df =1, 196) were computed to make individual comparisons among the six means. All the differences among the means for recall of the second input were significant (p ranging from .05 to .001). None of the differences among the means for recall of the first input approached significance. Comparing recall of the first and second input yielded significant ($p < .001$) differences in favor of the second input for Cond. O and F. The difference between the input conditions for Cond. S did not approach significance.

Numbers.—The main effects of input and output were significant: input, F (1, 98) = 7.61, $p < .01$; output, F (2, 196) = 21.03, $p <.001$. The Input × Output interaction was also significant, F (2, 196) = 3.30, $p < .05$. A series of individual contrasts, F (1, 96), showed that all the differences among the means for recall of the second input were significant ($p < .01$). Comparisons among the means for recall of the first input showed two significant differences: O > F, O > S ($p < .01$). Comparing recall of the first and second input yielded significant ($p < .01$) differences in favor of the second input for Cond. O and F. The differences between the input conditions for output "second" did not approach significance.

Conclusions

Experiment I confirmed the findings of the original study (Epstein, 1969, Exp. II) described in the introduction. Recall of lists presented second depended on the output specification in the same way as in the original experiment.

TABLE 2

MEAN NUMBER OF WORDS (Ws) AND NUMBERS (Ns) RECALLED FOR TWO INPUT ORDERS AND THREE RECALL OUTPUT SPECIFICATIONS: EXP.II

Input	Only		First		Second	
	M	*SD*	*M*	*SD*	*M*	*SD*
Word output						
W_1N_2	3.88	1.34	3.79	1.51	4.36	1.53
N_1W_2	5.68	1.20	5.04	1.24	4.46	1.74
Number output						
N_1W_2	3.98	1.42	3.40	1.15	2.81	.94
W_1N_2	4.44	1.47	3.46	1.35	3.04	.82

Note.—Total possible = 12.

The results were different for the list which was presented first. Recall of words was almost identical for the three output conditions. Recall of numbers was superior under Cond. O; recall for the two remaining conditions did not differ. In light of the results for words and the absence of any difference between numbers recalled first and second, the results for "numbers only" need to be checked.

EXPERIMENT II

Experiment II was like Exp. I except that 12-item lists were used instead of 8-item lists. The longer lists were introduced to determine whether the difference between the amounts recalled under Cond. O and F would be affected by increasing the number of items that S had to keep in circulation under Cond. F.

Method

Materials.—The lists used in Exp. I were supplemented by the addition of four items from Howes' (1966) and Battig and Spera's (1962) tables.

Subjects.—Fifty summer students of introductory psychology served as Ss. Twenty-five were assigned to each of the two input orders, i.e., N_1W_2 and W_1N_2.

Procedure.—The procedure was identical to Exp. I.

Results

Table 2 shows the mean number of words and numbers recalled. Separate analyses of variance were computed for word recall and number recall.

Words.—The main effect of input was significant, $F\ (1,\ 48) = 11.99,\ p < .01$. The main effect of output was not significant, but the Input × Output interaction was significant, $F\ (2,\ 96) = 6.86,\ p < .01$. The nature of the interaction can be discerned by examining Table 2. When words were the first input (W_1N_2), the mean recall did not differ from the three output conditions. This observation was confirmed by the results of individual contrasts, all of which were not significant. When words were put in second, recall varied with output condition: $O > F,\ F\ (1,\ 96) = 4.30,\ p < .05$; $F > S,\ F\ (1,\ 96) = 4.40,\ p < .05$; $O > S,\ F\ (1,\ 96) = 12.10,\ p < .01$.

Numbers.—There was only one significant main effect: output, $F\ (2,\ 96) = 20.72,\ p < .01$. Individual contrasts showed that for numbers put in second, $O > F\ (p < .05)$. For numbers put in first, $O > F\ (p < .01)$; $O > S\ (p < .01)$. Recall under Cond. F and S did not differ significantly.

Conclusion

The results of Exp. II confirm the results of Exp. I in all important details. For the second input, both words and numbers are recalled best under Cond. O.

EXPERIMENT III

On the "only" trial, *S* may use the full 60-sec. recall period to retrieve and reproduce items. This is not the case for the first output on the "first, then second" trial. On this trial, *S* must reserve the latter part of the recall period for output of the second list. Although informal observation of *S*'s performance suggested that *S*s used approximately the same amount of time to reproduce the "only" and "first" list, Exp. III was designed to control the time allowed for recall.

Method

Subjects.—The *S*s were 18 students of introductory psychology.

Procedure.—The materials and method of presentation were the same as in Cond. W_1N_2 of Exp. I. The instructions were modified by the addition of a statement describing the new feature of this experiment. The innovation in Exp. III was an extension of the recall period to 80 sec. and the introduction of a signal after 40 sec. of the recall period had elapsed. On the "only" trial, the signal was "stop"; on the "first, then second" trials, the signal was "shift to other list." This procedure equated the recall time allowed for the "only" and "first" recalled lists. One order of input, W_1N_2, was investigated.

TABLE 3

MEAN NUMBER OF WORDS AND NUMBERS WHEN FIXED TIME IS ALLOWED FOR ONLY AND FIRST RECALL: EXP. III

Item	Recall output					
	Only		First		Second	
	M	*SD*	*M*	*SD*	*M*	*SD*
Words	3.89	1.19	3.72	.89	3.76	.39
Numbers	3.47	1.09	2.86	.72	2.00	1.00
Combined	3.68	—	3.31	—	2.38	—

Results

The results are shown in Table 3. The results confirmed the findings of Exp. I. Analysis of variance showed a significant main effect of item, $F\ (1,\ 2) = 33.14,\ p < .001$, and a significant Item × Output Condition interaction, $F\ (2,\ 34) = 45.61$. The nature of the interaction effect can be discerned in Table 3. For the words, the first input, there was little difference among the three output conditions. For numbers, the second input, the number of items recalled was in the order $O > F > S$.

Conclusion

The superiority of "only" recall over "first" recall is obtained when time available for recall is equated. The results of Exp. I and II are not simply artifacts of uncontrolled differences in time allowed for recall.

EXPERIMENT IV

Experiment IV was designed to examine the effect of recall output specification, when the two lists are presented cotemporally. For this purpose, single lists were formed consisting of words and numbers.

Method

Materials.—The words and numbers used in Exp. I were used again. Each of the paired word and number lists was combined to form a single list of 16 items. The words and numbers were randomly arranged in the list with the restriction that four

TABLE 4

Mean Number of Words and Numbers Recalled from Randomly Mixed Lists for Three Recall Output Specifications: Exp. IV

Item	Recall output					
	Only		First		Second	
	M	*SD*	*M*	*SD*	*M*	*SD*
Words	3.69	.79	3.45	1.03	3.39	.66
Numbers	2.36	.82	2.36	1.17	1.77	.90
Combined	3.02	—	2.90	—	2.58	—

words and four numbers appeared in each half of the 16-item list.

Subjects.—The *S*s were 39 new students from the same population that provided *S*s for Exp. I.

Procedure.—The instructions were modified to describe the new composition of the learning lists. In all other respects, the procedure was the same as in Exp. I.

Results

Table 4 shows the mean number of words and numbers recalled for the three recall output specifications. The amount recalled was lower than in Exp. I. The mean number of words recalled under all recall conditions was 4.46 in Exp. I and 3.52 in Exp. IV; the mean number of numbers recalled was 2.60 in Exp. I and 2.16 in Exp. IV. The ratio of word to number recall was 1.71 in Exp. I and 1.61 in Exp. III.

Analysis of variance showed significant main effects for type of item, $F\ (1, 38) = 126.06$, $p \leq .001$, and output condition, $F\ (2, 76) = 5.18$, $p < .01$. The Item × Output interaction was of borderline significance, $F\ (2, 76) = 2.42$, $.05 > p < .10$. The main effect of type of item reflects the consistently superior recall of words compared with numbers. The main effect of output was examined by computing six individual contrasts. There were only three significant differences: words, O > S ($p < .05$); numbers, O > S, F > S ($p < .05$).

Conclusions

Recall under Cond. O was not superior to recall under Cond F. The main findings of Exp. I–III were not obtained with the mixed lists.

TABLE 5

Mean Number of Words and Numbers Recalled from Regularly Alternating Mixed Lists for Three Recall Output Specifications: Exp. V

Item	Recall output					
	Only		First		Second	
	M	*SD*	*M*	*SD*	*M*	*SD*
Words	3.70	.73	3.61	.81	3.05	.72
Numbers	2.26	1.02	2.02	1.08	1.78	.97
Combined	2.98	—	2.82	—	2.41	—

Experiment V

The mixed-list procedure of Exp. IV was employed to make presentation of the numbers and words cotemporal. However, the random composition of the list may also have made it difficult for *S* to sort the items into two separate memory compartments. This difficulty might have arisen because randomization of the list did not allow *S* to anticipate the category of the forthcoming item.

Experiment V was designed to retain the feature of cotemporality and, at the same time, to permit *S* to sort the items easily. This was accomplished by using mixed lists in which the numbers and words appeared in regular alternation. Since the instructions informed *S* that regular alternation would be followed in all lists, he could set himself in advance of each item for the appropriate storage sort.

Method

Materials.—The lists used in Exp. IV were rearranged to conform to the rule of regular alternation between words and numbers.

Subjects.—Forty-six students from the introductory psychology classes served as *S*s.[2]

Procedure.—The procedure was the same as in Exp. IV, with the exception of a modification in the instructions to inform *S* of the regular alternation between words and numbers.

---

[2] The first 26 *S*s were run at the University of Kansas by a research assistant in the summer of 1968. The last 20 *S*s were run at the University of Wisconsin by the author in the fall of 1968. The results for the two samples were highly similar. In fact, the means for two of the conditions were identical in the two samples.

Results

Table 5 shows the mean number recalled. Comparison of the means in Tables 4 and 5 shows that comparable levels of recall were obtained in Exp. IV and V. Apparently, regular alternation does not confer any advantages in recall. In Exp. V, as in Exp. IV, the level of recall was lower than in Exp. I, which presented two separate lists.

Analysis of variance showed main effects for type of item, $F\ (1,\ 45) = 60.46$, $p < .001$, and output condition, $F\ (2, 90) = 5.47$, $p < .01$. The interaction was not significant. The main effect of item is due to superior recall of words compared to numbers. Individual contrasts were computed to examine the effect of output condition. The following contrasts were significant at the .01 level: words, O > S, F > S; numbers O > S. Again as in Exp. IV, recall under Cond. O did not differ from recall under Cond. F. This was true both for numbers and words.

Conclusions

When the input condition makes the two sets of items cotemporal, there is no advantage for the "only" output over the "first" input. This is the case even if sorting of the items into two separate storage categories is facilitated by regularly alternating the items in the mixed list.

Discussion

Experiments I–III confirmed the observations of the earlier experiment by Epstein (1969). If two distinctive lists are successively presented for memorization, immediate recall is superior when recall of only one list is required compared to recall of the same list when *S* is also required to recall the other list. The experiments are not conclusive regarding the question of whether this superiority is restricted to the list that was presented second or whether this superiority will occur for the first input as well. Experiments IV and V show that for the material studied, separate, uninterrupted input of the two lists is a necessary condition of the effect.

The results seem similar to earlier findings by Anderson (1960) that poststimulus cuing for recall of only part of a list of 12 digits resulted in better performance than recall of the entire list. However, since Anderson did not report comparisons for the same set of digits when these were recalled alone and as the second or third set of the output, she did not make the observation reported here. In addition, the fact that superior recall of "only" lists was not obtained in Exp. IV and V indicates that the results in Exp. I–III are not simply instances of the effect demonstrated by Anderson.

Two processes may be jointly responsible for the superior recall of "only" lists: (*a*) The list which is not currently being recalled inhibits retrieval of the items from the presently recalled list. The degree of inhibition will be greater if the other list is being actively recirculated in preparation for subsequent recall. This is the case for the other list on the "first, then second" recall tests. On the "only" trial the excluded list is not rehearsed and inhibitory effects are less. (*b*) When the "only" signal is given, *S* may discard the other list from the short-term-memory store, thereby eliminating the potentially interfering material entirely. Preliminary observations in an experiment that compared recall of the excluded list following recall of the "only" list with recall of the same list when it was second on a "first, then second" trial showed superior recall in the latter case. This outcome is compatible with the first assumption since rehearsal of the second list would enhance recall of that list. The fact that recall of the excluded list after the "only" list did not diminish to zero suggests that the other list is not completely discarded on the "only" trials.

The fact that the superiority of Cond. O does not appear with the mixed lists of Exp. IV and V requires explantion. One possibility is that when the input is mixed and *S* is asked to recall only one type of item, the excluded items, nevertheless, intrude and inhibit retrieval of the prescribed items. The result is to make Cond. O and F functionally more similar and to reduce the difference between the number of prescribed items recalled under these conditions. When the two lists are presented separately, as in Exp. I–III, *S* is more successful in suppressing the excluded list under Cond. O. Items that are grouped together in input can be excluded as a group during output.

The experiments add to the growing body of data (e.g., Tulving, 1967; Tulving & Pearlstone, 1966), showing that in addition to processes during acquisition, there are processes during retrieval that affect recall.

REFERENCES

Anderson, N. S. Poststimulus cuing in immediate recall. *Journal of Experimental Psychology*, 1960, **60**, 216–221.

BATTIG, W., & SPERA, A. J. Rated association values of numbers from 0-100. *Journal of Verbal Learning and Verbal Behavior*, 1962, 1, 200–202.

EPSTEIN, W. Recall of word lists following learning of sentences and of anomalous and random strings. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 20–25.

HOWES, D. A word count of spoken English. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 572–604.

TULVING, E. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 175–184.

TULVING, E., & PEARLSTONE, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received February 13, 1969)

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Category Clustering for Immediate and Delayed Recall as a Function of Recall Cue Information and Response Dominance Variability: Robert L. Hudson*: Department of Psychology, Mississippi State University, State College, Mississippi 39762.

Evaluation of a Model Relating Thorndike-Lorge Frequency and *m* to Learning: Vito Modigliani and Eli Saltz*: Department of Psychology, Wayne State University, Detroit, Michigan 48202.

Free Recall of Numbers with High- and Low-Rated Association Values: Stefan Slak*: Department of Psychology, University of Toledo, Toledo, Ohio 43606.

Order of Report and Coding in Memory: Harvey A. Taub* and Richard A. Monty: Veterans Administration Hospital, Irving Avenue and University Place, Syracuse, New York 13210.

Eye Tracking of Self-Moved Targets: Role of Efference: Martin J. Steinbach*: Department of Psychology, York University, 4700 Keele Street, Downsview, Toronto 12, Ontario, Canada.

Idiodynamic and Popular Sets in Word Associations of College Students: Ronald Newman and Theodore R. Dixon*: Department of Psychology, Texas Christian University, Fort Worth, Texas 76129.

Transformations on Sets in Short-Term Memory: Set Size Reduction by Deletion: Donald V. DeRosa*: Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43402.

Why Are Rhymes Easy to Learn?: Gordon H. Bower* and Laura S. Bolton: Department of Psychology, Stanford University, Stanford, California 94305.

Effects of Facilitory and Inhibitory Sets on GSR Conditioning and Extinction: Michael E. Dawson* and Paul Reardon: Department of Psychology, California State College at Los Angeles, 5151 State College Drive, Los Angeles, California 90032.

Observing Responses, Attention, and the Overtraining Reversal Effect: Peter D. Eimas*: Walter S. Hunter Laboratory of Psychology, Brown University, Providence, Rhode Island 02912.

Proactive Inhibition and Undetected Retention Interval Rehearsal: John P. Houston*: Department of Psychology, University of California, Los Angeles, California 90024.

Differential Cue Habit Strength as a Determinant of Attention: Joseph C. Campione* and Prudence C. Wentworth: Department of Psychology, Developmental Psychology Laboratory, University of Washington, Seattle, Washington 98105.

Combined Effects of Fear and Frustration on Acquisition of a Hurdle-Jump Response: Helen Bohmer Daly*: Department of Psychology, State University of New York College at Oswego, Oswego, New York 13126.

Effect of A-B Overtraining in A-Br: Carlton T. James* and James G. Greeno: Department of Psychology, University of Texas, Austin, Texas 78712.

Training of Short-Term Memory: Frank Restle*: Department of Psychology, Indiana University, Psychology Building, Bloomington, Indiana 47401.

---

* Asterisk indicates author for whom address is given.

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 175–182

SHORT-TERM RETENTION: PREPARATORY SET AS COVERT REHEARSAL[1]

THEODORE J. DOLL[2]

Mental Health Research Institute, University of Michigan

It is shown that preparatory processing produced by advance partially informative cueing in reaction time (RT) tasks which employ verbal stimuli can be viewed as rehearsal in short-term-memory tasks. Experiment I showed that the advance-cueing RT procedure produces decreased mean RT to cued stimuli and increased mean RT to noncued stimuli, suggesting that the cues evoke enhanced preparatory processing of cued stimuli and attenuated preparatory processing of noncued stimuli. Experiment II demonstrated that enhanced preparatory processing, like rehearsal, facilitates short-term recall and that attenuated preparatory processing, like attenuation of rehearsal by interpolated activity, impairs recall. The implications of the findings for the necessary and sufficient conditions for rehearsal, the selectivity of rehearsal, and the relationship of rehearsal to the accessibility of information in long-term memory are discussed.

Covert rehearsal has become an important explanatory concept in the psychological literature on short-term retention (Brown, 1958; Conrad, 1960; Hellyer, 1962; Murdock, 1961; Waugh & Norman, 1965). However, little is known about the various circumstances in which rehearsal occurs and the conditions which determine whether or not rehearsal occurs in any given situation.

The present study shows that preparatory set or preparatory processing produced by advance partially informative cueing in reaction time (RT) tasks which employ verbal stimuli can be viewed as rehearsal in short-term-memory tasks. In addition to demonstrating the relationship of processes in the two paradigms, the study has implications for the necessary and sufficient conditions for rehearsal, the selectivity of rehearsal, and relationship of rehearsal to the accessibility of information stored in long-term memory.

The purpose of Exp. I was to show that advance partially informative cueing in an RT task produces differential preparatory processing of the cued and noncued stimuli. Davis (1964) and Shaffer (1966) have shown that advance cues which signal a reduction in the number of alternative stimuli to which *S* may be required to react (independent of response uncertainty) produce a reduction in mean RT to the cued stimuli. This finding suggests that the cues evoke enhanced preparatory processing of the cued stimuli.

In Exp. I the advance cues signaled a change in the probabilities (always nonzero) of the stimuli rather than a reduction in the number of alternative stimuli. The hypothesis tested was that cueing produces an increase in mean RT to noncued stimuli as well as a reduction in mean RT to cued stimuli. Such a finding would suggest that advanced cues produce attenuated preparatory processing of the noncued stimuli as well as enhanced preparatory processing of the cued stimuli.

The purpose of the second experiment was to determine whether the preparatory process controlled by the advance-cueing RT procedure can be viewed as rehearsal.

---

[1] This paper is based on a dissertation submitted to the Kent State University Graduate School in partial fulfillment of the requirements for the PhD degree. The author is grateful to Robert E. Morin for his advice in the design and writing of the dissertation and to Sylvan Kornblum, Irwin Pollack, and Jane and William Siegel for their helpful criticism in the writing of this paper. The research was supported by National Science Foundation Grant GB-4343X awarded to Kent State University and was conducted while the author held a United States Public Health Service predoctoral fellowship (5-F1-MH-36,830-02).

[2] Requests for reprints should be sent to Theodore J. Doll, who is now at the State University of New York at Stony Brook, Department of Psychology, Stony Brook, New York 11790.

TABLE 1

STIMULI, CONDITIONS, AND EXAMPLES OF TRIALS IN THE RT TASK (EXP. I)

Type of probe	Trial events				P (any item from probed subset/cue)
	Starting signal	Preprobe interval	Probe	Correct response	
Cued	2	Enhanced preparation for QUR	QX	Left	.80
Miscued	2	Enhanced preparation for QUR	BX	Left	.10
Uncued	Ready	Uniform preparation for all subsets	XP	Right	.33

Note.—The critical set consisted of the subsets BDP, QUR, and SFL, cued by the digits 1, 2, and 3, respectively. The noncritical set consisted of the letters T V Z C G O M N X.

The basis for the inference about the relationship of the processes was the effect of the preparatory process on short-term recall. If preparatory processing can be viewed as rehearsal, then enhanced and attenuated preparatory processing should produce the effects on short-term recall that are customarily attributed to rehearsal and attenuation of rehearsal by interpolated activity, respectively. That is, enhanced preparatory processing should facilitate recall and attenuated preparatory processing should impair recall.

EXPERIMENT I

Method

Subjects.—The *S*s were selected on the basis of their performance in free-learning a set of nine critical letters (see Table 1). Of 20 *S*s, 12 who learned the critical set to a criterion of nine consecutive correct responses in 54 or fewer trials participated in the balance of Exp. I. All *S*s were male undergraduates at Kent State University. The *S*s were paid and received credits toward a course requirement for their participation.

Apparatus.—The *S*s were trained and tested individually in a sound-deadened chamber. The *S* was seated approximately 3 ft. in front of an eye-level array of three alphanumeric display cells (Industrial Electronic Engineers). The entire array was 3.5 in. wide and 6.5 in. high; stimuli were 1.5 in. high and 1.0 in. wide. Stimulus sequences were preprogrammed and automatically executed. Reaction time was measured with millisecond accuracy by a Hewlett-Packard electronic counter. The manipulandum was a small lever attached to a spring-loaded, center-positioning, double throw switch, placed within easy reach of *S*'s preferred hand. Movement of the lever to the left or right terminated the probe stimuli.

Stimuli.—The critical and noncritical items shown in Table 1 were so chosen in order to systematically vary phonemic and visual confusability. It was of interest to determine in what form information is encoded for preparatory processing. The manipulations of phonemic and visual confusability were constrained by the primary design considerations. As a result, no evidence was obtained for either phonemic or visual encoding. Further information regarding the selection of the stimuli and the confusability manipulations was reported by Doll (1968).

Procedure.—Each *S* participated in two 1-hr. sessions. In each session *S* first performed a free-learning task, then a paired-associates learning task, and finally the RT task.

On each trial of the free-learning task the nine critical letters were presented at a 2/sec rate in a new random order. The word "ready" appeared 3 sec. after the offset of the last letter as a signal to *S* to recall the letters in any order. After 9 sec., the ready signal terminated and was followed by another presentation of the nine letters. The task was terminated after *S* met the criterion or after 54 trials, whichever occurred first. The *S* was dismissed if he had not met the criterion by the end of Session 1.

In the paired-associates learning task, the digits 1, 2, and 3 served as stimuli, and the responses consisted of mutually exclusive subsets of the previously learned list of nine letters (see Table 1). On each trial the digit was displayed for 3 sec., during which *S* was required to say aloud, and in any order, the appropriate subset of letters. The correct letters then appeared at a 2/sec rate and were followed by a 3-sec. intertrial interval. The task was terminated when *S* had made 12 consecutive correct anticipations.

The procedure for the RT task is illustrated in Table 1. A pair of probe letters was presented on each trial. The *S* was required to indicate which letter in the pair belonged to the critical set by pushing the response lever to the side which corresponded to its position in the display. The *S* was instructed to respond as quickly as possible without making errors. The intertrial interval was 3 sec.

Each probe included one critical and one noncritical letter. The *S* was not acquainted with the noncritical set. Each trial was initiated by a .5-sec. starting signal which could be either a cue, i.e., the digit 1, 2, or 3, or the word "ready." A 3-sec. preprobe interval followed, during which the display was blank. The *S* was instructed that the cues would provide partial advance information about which critical letter would appear in the probe. The *S* was not told to rehearse, or in any way prepare for, the cued subset.

When the starting signal was a digit, the probe was correctly cued with probability .8 (cued probes), i.e., the critical letter in the probe was drawn from the cued subset with $p = .8$. With probability .2, the critical letter in the probe was drawn from the remaining two (noncued) subsets with equal likelihood (miscued probes). When the starting signal was "ready," the critical item in the probe was drawn equally often from the three subsets (uncued probes).

The critical probe letter appeared equiprobably on the left and right in all conditions; hence stimulus uncertainty varied independently of response uncertainty. The nine critical letters appeared in the probe equally often in the cued, miscued, and uncued conditions. Therefore, prior to the presentation of the starting signal, stimulus uncertainty was the same in all conditions.

In the first session, *S* performed one practice block of 108 trials of the RT task. These data were not analyzed. In Session 2, *S* performed three blocks of 108 trials each. Each block was made up of a different random sequence of 14 types of trials defined by the combination of three variables, type of probe, cue, and response. Cues 1, 2, and 3 appeared equally often with cued probes and equally often with miscued probes. The word "ready" appeared exclusively with uncued probes. The critical and noncritical letters for each probe were chosen randomly from the appropriate critical subset and the noncritical set, respectively. After each block in the second session, *E* advised *S* of the number of errors he had committed and the approximate range of his RTs. At the close of the experiment, *S* answered a questionnaire which elicited an introspective account of his performance.

Results and Discussion

Mean RTs for correct responses to cued, miscued, and uncued probes were 599, 719, and 647 msec., respectively. These results support the interpretation that partially informative cues evoke enhanced preparatory processing of cued stimuli and attenuated preparatory processing of noncued stimuli. The data are means of 12 individual-*S* means, each of which was based on at least 54 trials. An analysis of variance which included correct response and type of probe as within-*S* variables revealed no significant effect of the former variable, $F\ (1, 11) = .51$, $p > .05$, and confirmed the previous results. Mean RT for correct responses to uncued probes differed significantly from that for cued probes, $F\ (1, 11) = 17.70$, $p < .01$, and from that for miscued probes, $F\ (1, 11) = 18.93$, $p < .01$. The variances of RT suggest that these differences are not due to selective retrieval processes which occur after the probe is presented. Suppose that: (*a*) the cue determined the order in which the subsets were searched, i.e., the cued subset was searched first, (*b*) the search order was random or haphazard when the starting signal was uninformative, and (*c*) the search terminated as soon as the subset which contained the searched-for item was encountered. This interpretation accounts for the observed differences in mean RT, but predicts greater variability in RT for uncued probes (random search order) than for cued probes (cue-controlled search order). In fact, the mean variance of RT for correct responses was smaller for uncued probes (26,190 msec.^2) than for cued probes (31,762 msec.^2). The mean variance for miscued probes was 47,077 msec^2. These data are means of 12 individual-*S* variances.

The particular cue presented had relatively little effect on RT. Mean RTs for correct responses to cued probes were 584, 619, and 596 msec. for Cues 1, 2, and 3, respectively. Mean RTs for correct responses to miscued probes were 707, 728, and 725 msec. for Cues 1, 2, and 3 respectively. Again these are means of 12 individual-*S* means. For cued probes each individual-*S* mean was based on 72 trials. The individual-*S* means for miscued probes were each based on 18 trials. The effect of cue was significant for cued probes, $F\ (2, 22) = 5.15$, $p < .05$, but not for miscued probes, $F\ (2, 22) = .60$, $p > .05$.

The mean proportions of errors for responses to cued, miscued, and uncued probes were .031, .061, and .061, respectively. Error rate was also relatively independent of the particular cue presented. For cued probes, the mean proportions of errors for Cues 1, 2, and 3 were .026, .039, and .028, respectively. The mean proportions of errors for miscued probes were .050, .074, and .060 for Cues 1, 2, and 3, respectively. The data for errors are means of 12 individual-*S* proportions.

The responses to the questionnaire strengthened the conclusion that the cues

TABLE 2

CONDITIONS AND EXAMPLES OF TRIALS IN THE EXPERIMENTAL TASK (EXP. II)

Type of probe	Subset probed	Trial events					
		Critical set		Retention interval and starting signal[a]	RT probe	Correct response	Free recall
		Subset 1	Subset 2				
C	1	L W Q	P H X	1	Q K	Left	*S* attempts to recall the critical set
C	2	L Z D	S G N	2	J G	Right	
UC	1	B R D	Z T C	3	B N	Left	
UC	2	H P N	T J L	3	Q T	Right	

Note.—C represents cued probes; UC represents uncued probes.

[a] Retention intervals of 2, 4, 6, and 8 sec. occurred equiprobably in random order, and each was followed by a .5-sec. warning before the probe.

controlled the preparatory process. The *S*s were asked, "In the reaction time part of the experiment, did the number (1, 2, or 3) which sometimes preceded the pairs of letters make your task easier or harder?" and "Why do you feel that the numbers made your task either easier or harder?" Nine *S*s replied that the cues made the RT task easier because they could "concentrate on," "think of," etc., only three rather than nine critical letters. The other three *S*s remarked that the cues made the RT task harder because they tended to concentrate on members of the cued subset, and this "threw them off," "surprised them," etc., when the critical letter was not a member of that subset (miscued probes). To the third question, "What did you think about after the number came on, and before the pair of letters appeared?" nine *S*s said they thought of the letters associated with the digit, one *S* said he thought of all nine critical letters, and two *S*s' statements were not clear.

EXPERIMENT II

Method

Subjects and apparatus.—The *S*s were selected on the basis of their performance in a practice task. Of 17 *S*s, 13 met the selection criteria. Of these 13 *S*s, 2 were run in a pilot version of the experiment and 2 could not be run because of an apparatus failure, leaving 9 *S*s in the experiment. All *S*s were male undergraduates at Kent State University. The *S*s were paid and received credits toward a course requirement for their participation. With one addition, the apparatus was the same as that used in Exp. I. A small green signal lamp mounted near the response lever provided the recall signal.

Procedure.—Each *S* participated in two sessions. The first session required about 30 min., during which *S* performed 32 trials of the practice task. The data from the practice task were not analyzed. Session 2 required about 90 min. The *S* performed six blocks of 32 trials of the experimental task. The order of presentation of blocks was randomized across *S*s.

The procedure for the experimental task is outlined in Table 2. Each trial began with the presentation of six new critical consonants at the rate of 2/sec. The sets of critical consonants were taken from Hintzman's (1966) table of random letters. Sets which contained a repetition or a mnemonic series, e.g., NBC, were excluded. The first three letters of each set were called Subset 1; the last three were called Subset 2.

Immediately after the presentation of the last critical consonant, a starting signal appeared and remained throughout a retention interval which was equiprobably 2, 4, 6, and 8 sec. The starting signal could be either a cue, i.e., the digit 1 or 2, or the digit 3, which served the same function as did the word "ready" in Exp. I. The starting signal was followed by a .5-sec. warning interval, during which the display was blank. The warning interval was designed to minimize effects of time uncertainty on RT.

Two probe letters, one critical and one noncritical, followed the warning interval. When the starting signal was the digit 1 or 2, the probe was correctly cued with probability unity (cued probes), i.e., the critical letter in the probe was randomly drawn from the subset of the same number. When the starting signal was the digit 3, the critical item in the probe was drawn randomly and equiprobably from the two subsets (uncued probes). The noncritical letter was chosen randomly from among the 14 consonants which had not appeared in the critical set.

The *S* was instructed to push the lever to the side which corresponded to the position of the critical letter in the probe display as quickly as possible without making errors. After the RT response, the green signal lamp glowed for 6 sec. The *S* was required to recall as many critical letters as possible in any order while the green lamp remained on. An intertrial interval of 2 sec. followed recall.

In the experimental task there were 32 equiprobable types of trials defined by the combination of two levels of type of probe, two levels of subset probed, two levels of correct response, and four levels of retention interval. One instance of each trial type appeared in a different random position in each block. The practice task differed from the experimental task only in that there was one level of type of probe (uncued). Two exemplars of each of 16 trial types appeared in random order in the practice task. At the end of Session 1, *S* was dismissed from the experiment if his performance in the practice task did not meet the following criteria: (*a*) eight or more complete recalls, (*b*) eight or fewer RT errors, and (*c*) eight or fewer RTs equal to or greater than 2 sec.

The retention interval was randomly varied in order to determine whether the preparatory process occupied its full duration. If the retention interval had been constant over blocks of trials, it could be argued that preparation was delayed until just prior to the probe. With a randomly varied retention interval, such a strategy would result in little or no preparation for probes which followed short retention intervals. Therefore, the difference in RT for cued and uncued probes would be smaller at short retention intervals than at long retention intervals. On the other hand, if a full state of preparation was achieved early and was preserved throughout the retention interval, the difference in RT for cued and uncued probes should remain constant over the length of the retention interval.

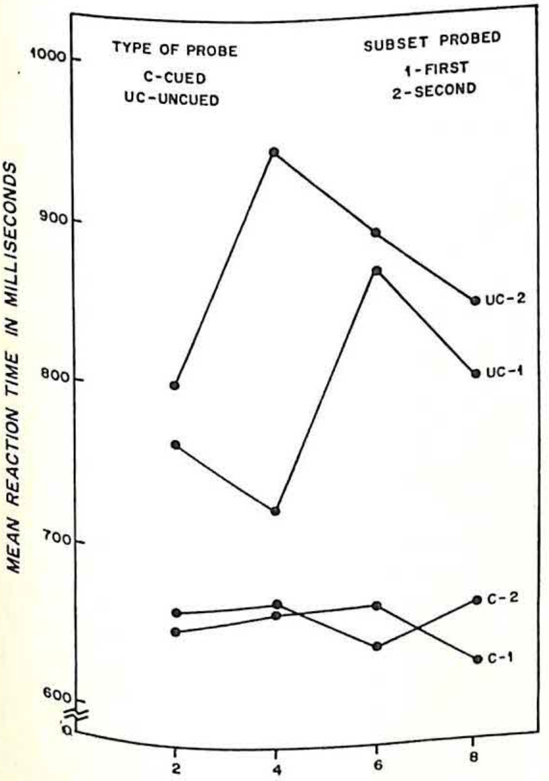


FIG. 1. Mean RT as a function of type of probe, subset probed, and retention interval (Exp. II).

In the second session, *S* was instructed that the digits 1 and 2 would give him advance information about the subset from which the critical letter in the probe would be drawn. He was not told to rehearse, or in any way prepare for, the cued items. The *S* was informed that he would receive 1¢ additional pay for each correct response with an RT less than .600 sec. and 1¢ less pay for each incorrect RT response.

After each block of the experimental task, *S* was informed of the number of correct responses with RTs less than .600 sec. and the number of errors that he had made in the block. Both the incentive for fast and accurate RT responses and this feedback were used to emphasize the importance of the RT task relative to recall. In addition, *S* was told that recall was required only "to make sure you were not guessing" which RT response was correct. (These procedures were instituted after data from pilot *S*s revealed no difference in mean RT for cued and uncued probes. It was thought that the requirement that *S* recall all six letters produced uniform preparatory processing of all the letters and that this would be avoided if recall appeared to be incidental.) After *S* completed the experiment, he was asked to fill out a questionnaire.

Results and Discussion

In Fig. 1, mean RT for correct responses is plotted as a function of retention interval, with type of probe and subset probed as parameters. Each point is the mean of nine individual-*S* means, each of which was based on 12 trials. An analysis of variance which included type of probe, subset probed, and retention interval as within-*S* factors revealed only three significant effects. Mean RT for correct responses to cued probes was significantly faster than that for uncued probes, $F\ (1, 8) = 37.25$, $p < .001$, suggesting that the cues controlled the preparatory process. The interactions of type of probe by the three orthogonal trends with retention interval (linear, quadratic, and cubic) were not significant, $F\ (1, 8) = 3.09$, 1.74, and .24, respectively, $p > .05$ in all cases, suggesting that a full state of preparation was achieved early and was maintained throughout the retention interval. Mean RT was significantly faster when Subset 1 was probed than when Subset 2 was probed, $F(1, 8) = 6.82$, $p < .05$. This result suggests that the first-presented subset underwent greater preparatory processing than the second-presented subset. The only other significant effect was the interaction of subset probed by the cubic trend with retention interval, $F(1, 8) = 6.01$, $p < .05$.

Each trial was assigned two recall scores: the number of letters of Subset 1 recalled and the number of letters of Subset 2 recalled. A letter was counted as correctly recalled if it appeared anywhere in *S*'s response, even if accompanied by erroneous letters. Recall data for the first and second subsets are displayed in Fig. 2 and 3, respectively. The curves labeled C-1 are the data for trials on which Subset 1 was cued and subsequently probed. The C-2 curves are the data for trials on which Subset 2 was cued and probed. The curves labeled UC-1 and UC-2 are the data for the control conditions for C-1 and C-2, respectively. In Cond. UC-1, neither subset was cued, but like Cond. C-1, Subset 1 was probed. In Cond. UC-2, neither subset was cued, but as in Cond. C-2, Subset 2 was probed.

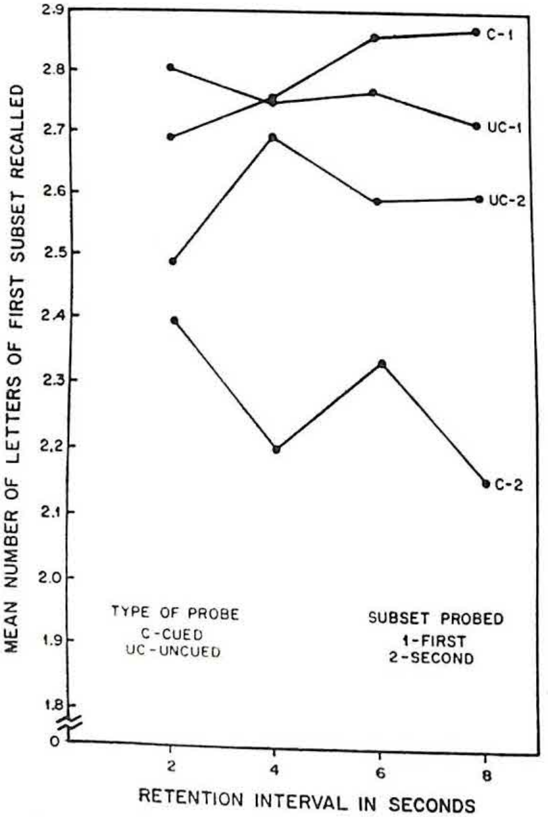


FIG. 2. Mean number of letters of the first subset recalled as a function of type of probe, subset probed, and retention interval (Exp. II).

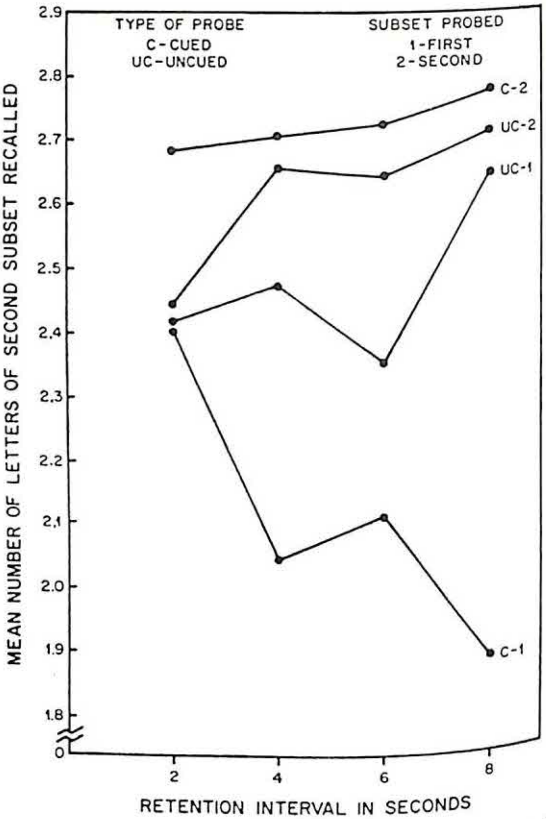


FIG. 3. Mean number of letters of the second subset recalled as a function of type of probe, subset probed, and retention interval (Exp. II).

The overall effect of cueing was determined by comparing the average vertical displacements of Curves C-1 and C-2 with those of the control curves. First consider the recall of Subset 1 (see Fig. 2). The first subset was not recalled significantly better when it was cued than in the control condition—comparing Curves C-1 and UC-1 in Fig. 2 yields $F(1, 8) = .67$, $p > .05$. However, when the other subset was cued, recall of Subset 1 was significantly poorer than in the control condition—comparing C-2 and UC-2 yields $F(1, 8) = 17.18$, $p < .01$. Similar results were obtained for Subset 2 (see Fig. 3). When Subset 2 was cued, it was not recalled significantly better than in the control condition—comparing C-2 and UC-2 yields F (1, 8) = 4.23, $p > .05$. However, when the other subset was cued, Subset 2 was recalled significantly more poorly than in the control condition—comparing C-1 and UC-1 yields F (1, 8) = 7.88, $p < .05$.

The linear trends over the retention interval were also compared. Consider the data for Subset 1 in Fig. 2. When Subset 1 was cued, its recall improved significantly in relation to the control—C-1 vs. UC-1 yields F (1, 8) = 8.00, $p < .05$. When the other subset was cued, recall of Subset 1 did not significantly deteriorate relative to the control—C-2 vs. UC-2 yields F (1, 8) = 2.04, $p > .05$. The converse results were obtained for Subset 2 (see Fig. 3). When Subset 2 was cued, its recall did not improve significantly relative to its control—comparing C-2 and UC-2, F (1, 8) = 1.52, $p > .05$. When the other subset was cued, recall of Subset 2 declined significantly in relation to its control—comparing C-1 and UC-1, F (1, 8) = 15.06, $p < .01$.

To summarize, cueing of a subset of items, i.e., enhanced preparatory processing, did not significantly facilitate mean recall over all retention intervals. This may be due to the fact that recall was already nearly perfect in the absence of cueing (i.e., in the control conditions). Enhanced preparatory processing did, however, produce a significant increase in recall of Subset 1 with the length of the retention interval. Cueing produced an overall decrement in the recall of the noncued subset, i.e., the subset for which preparatory processing was attenuated. In addition, attenuated preparatory processing resulted in a marked deterioration in recall of Subset 2 and a noticeable but nonsignificant deterioration in recall of Subset 1 with the length of the retention interval. The greater rate of forgetting for Subset 2 than for Subset 1 is consistent with the RT data which suggest greater preparatory processing of the first-presented subset. The fact that the effects of cueing on recall depend on the length of the retention interval strongly suggests that the process controlled by the cues takes place during the retention interval rather than, e.g., at the time of retrieval.

An overall analysis of recall data included four within-S factors: (*a*) type of probe, (*b*) subset probed, (*c*) subset, i.e., whether the data were for Subset 1 or 2, and (*d*) retention interval. Mean recall was greater after uncued probes than after cued probes, F (1, 8) = 9.25, $p < .05$, indicating that uniform preparation for both subsets produced better recall than concentrated preparation for a single subset. The opposite result might have been obtained if the critical set had been longer than the "memory span." Mean recall was better for Subset 1 than Subset 2, F (1, 8) = 6.51, $p < .05$. This primacy effect is consistent with the aforementioned evidence that the first-presented subset underwent greater preparatory processing. The Subset Probed × Subset interaction was significant, F (1, 8) = 67.31, $p < .001$, indicating that recall was best for the subset from which the critical letter in the probe was drawn. The interactions of Type of Probe × Subset Probed × Subset, F (1, 8) = 14.26, $p < .01$, and Type of Probe × Subset Probed × Subset × Linear Component of Retention Interval, F (1, 8) = 13.35, $p < .01$, reflect the same sources of variation encountered in the comparisons of average vertical displacements and linear trends over the retention interval discussed previously. The analysis also revealed five statistically significant but relatively unimportant interactions: Type of Probe × Retention Interval Linear, F (1, 8) = 6.08, $p < .05$; Type of Probe × Retention Interval Cubic, F (1, 8) = 69.30, $p < .001$; Type of Probe × Subset × Retention Interval Linear, F (1, 8) = 6.45, $p < .05$; Subset Probed × Subset × Retention Interval Linear, F (1, 8) = 6.52, $p < .05$; and Type of Probe × Subset Probed × Subset × Retention Interval Cubic, F (1, 8) = 6.34, $p < .05$.

As was the case in Exp. I, the mean proportion of errors for the RT response was greater for uncued probes than cued probes (.108 and .069, respectively). The questionnaire provided introspective support for the conclusion that the cues controlled the preparatory process. To the first two questions, "Did the numbers 1 and 2 which sometimes preceded the pairs of letters make your decisions easier or harder?" and "Why do you feel that the 1 and 2 made your decision easier or harder?" all nine *S*s answered "easier" because they could "concentrate on," "think of," etc., one subset rather than two. In response to the third question, "What did you think about while the numbers 1, 2, or 3 were on and before you made your decision?" three *S*s said that they thought of both subsets when the 3 appeared, but only the cued subset when a 1 or 2 appeared. Six *S*s' responses to the third question were not clear.

CONCLUSIONS

The RT data of Exp. I and II support the conclusion that advance partially informative cues evoke differential preparatory processing of cued and noncued stimuli. The recall data of Exp. II strongly suggest that preparatory processing and rehearsal are the same process. The deleterious effect of attenuated preparatory processing is comparable to the disruptive effect of interpolated activity (Brown, 1958; Peterson & Peterson, 1959). The facilitative effect of enhanced preparatory processing and the differential rates of forgetting for the first- and second-presented subsets are consistent with evidence reported by Crawford, Hunt, and Peak (1966), Hellyer (1962), Peterson and Peterson (1959), and Sanders (1961) that rehearsal strengthens, rather than merely renews, material being retained. Amounts and rates of forgetting in this and previous studies are not easily compared because free recall, rather than ordered recall, was used here. Also, the dependent variable used in this research (number of letters recalled) probably resulted in less apparent forgetting than the measure used in previous studies (percentage of correct ordered recalls) because the former is not sensitive to reciprocal intrusions of correct elements and inclusions of erroneous elements.

The results of Exp. II also suggest that rehearsal can be selective, e.g., the advance-cueing RT procedure can be used to control which of a set of stimuli are most rehearsed. The fact that the advance-cueing RT procedure produced impaired recall of noncued stimuli as well as improved recall of cued stimuli suggests that the rehearsal of one set of stimuli can interfere with the rehearsal of other stimuli. It appears, therefore, that overt interpolated activity is not a necessary condition for attenuation of rehearsal and that an unfilled interval, i.e., the absence of overt activity, is not a sufficient condition for (nonattenuated) rehearsal.

The findings of these experiments support Sternberg's (1968) conclusion that one function of rehearsal is to make information in long-term memory more accessible. The RT data of Exp. I suggest that the advance-cueing RT procedure affects the accessibility of information in long-term memory. The results of Exp. II strongly suggest that the process which mediates these changes in accessibility is rehearsal.

REFERENCES

BROWN, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

CONRAD, R. Very brief delays of immediate recall. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 45–47.

CRAWFORD, J., HUNT, E., & PEAK, G. Inverse forgetting in short-term memory. *Journal of Experimental Psychology*, 1966, **72**, 415–422.

DAVIS, R. The combination of information from different sources. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 332–339.

DOLL, T. J. Short-term retention: Covert rehearsal controlled by post-stimulus cues. (Doctoral dissertation, Kent State University) Ann Arbor, Mich.: University Microfilms, 1968. No. 69-9546.

HELLYER, S. Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology*, 1962, **64**, 650.

HINTZMAN, D. L. Tables of random letters. *Psychonomic Science*, 1966, **5**, 253.

MURDOCK, B. B. The retention of individual items. *Journal of Experimental Psychology*, 1961, **62**, 618–625.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

SANDERS, A. F. Rehearsal and recall in immediate memory. *Ergonomics*, 1961, **4**, 25–34.

SHAFFER, L. H. Some effects of partial advance information on choice reaction with fixed or variable S-R mapping. *Journal of Experimental Psychology*, 1966, **72**, 541–545.

STERNBERG, S. Inferring the organization of cognitive events from reaction-time: Some scanning processes in active memory. Paper presented at the Symposium on Research in Cognitive Psychology, City University of New York, June 1968.

WAUGH, N. C., & NORMAN, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

(Received February 17, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 183–184

SUPPLEMENTARY REPORTS

FRUSTRATIVE REWARD[1]

K. T. STRONGMAN[2] AND P. E. WOOKEY
University of Exeter

Four groups of rats received 80 training trials in an L-shaped double runway. Groups E(0) and E(2) received one food pellet in Goal Box 1 (GB_1) and two pellets in GB_2. During 40 subsequent test trials, E(0) experienced nonreward in GB_1 on 50% of the occasions, whereas E(2) experienced increased reward (two pellets) 50% of the time. Groups C(0) and C(2) served as controls for response suppression. Running times in Runway 2 demonstrated the usual frustration effect (FE), but also provided evidence for a "frustrative reward" effect in that E(2) showed an FE on *normal* test trials after brief experience with increased reward on experimental test trials.

While considerable emphasis has been placed on the effects of reduced reward in the double-runway situation (Amsel & Roussel, 1952; Wagner, 1959), other variations in reward have been generally ignored. The present study compared the effect of increased reward on the well-established frustration effect (FE).

Method.—Twenty male hooded rats of the Sprague-Dawley strain, aged about 150 days at the beginning of training, served as *S*s and were randomly divided into four groups of 5. During the experiment, 1 *S* died, and the data presented are from the remaining 19.

An L-shaped double runway constructed of plywood was used. The start box (SB), Runway 1 (A_1), and Goal Box 1 (GB_1) were painted mat white and had a wooden floor, while Runway 2 (A_2) and Goal Box 2 (GB_2) were mat black with a floor covering of perforated zinc. All wooden interior surfaces were coated with clear polyurethane varnish. The start and goal boxes were 12 in. long, and the runways were 60 in. long with a 3½-in. height and a 3-in. width throughout, except for GB_2 which was 5 in. wide. The entire apparatus was covered with hinged sections of clear Plexiglas and stood 18 in. above the floor. In GB_1 the food cup was a small aluminum tray located at floor level on the left of the box, whereas in GB_2 a larger tray was mounted 1 in. from the floor on the end wall of the box.

Externally manipulatable sliding doors separated the SB and GBs from the runway. Each door was made of colored Plexiglas, the sides corresponding in shade to that part of the maze which they faced. In A_2, a photocell was situated in the wall 1 in. above the floor and 12 in. from GB_1. Interruptions of this photocell beam activated a Venner millisecond timer, which stopped when a second photocell beam located a further 36 in. along A_2 was broken.

The experiment lasted for 50 days, this period being divided into three sections. The *S*s were not run on Days 11, 14, 19, 22, 23, 25, and 40 due to apparatus difficulties. The *S*s were trained in blocks of four, a block containing one *S* from each group.

Days 1–6 were spent in habituating *S*s to handling and reducing their weight to approximately 80% of free feed level, allowing 3 hr. of feeding following the time at which their block would later be run. Subsequently, this 21-hr. regime was continued with the addition of a growth allowance of 5% of prior free feed weight. On Days 7 and 8, *S*s explored the runway with all doors open for an overall total of 24 min. During these times, the goal boxes contained no food. At 10-min. intervals on Days 9 and 10, each *S* was placed four times into a separate goal box, similar in construction to those of the runway. On Day 9, the food tray contained a pellet of *S*s' normal diet together with a number of pellets of the type later used as reward; on Day 10, only the latter were present.

During experimental training (Days 12–39), the procedure for an individual trial was as follows: *S* placed in SB; start door opened when *S* faced it; *S* traversed A_1, entered GB_1, receiving food or not according to the experimental plan; retrace door of GB_1 closed; 10 sec. after entry into GB_1, start door opened, allowing access to A_2; *S* traversed A_2, entered GB_2, and received food; retrace door of GB_2 closed, and *S* removed as soon as food consumed. The four groups received .45-mg. dextrose Noyes pellets according to the following schema: E(0)—one pellet in GB_1, two pellets in GB_2; E(2)—one pellet in GB_1, two pellets in GB_2; C(0)—zero pellets in GB_1, two pellets in GB_2; C(2)—two pellets in GB_1, two pellets in GB_2. During this training period, each *S* received 80 trials: 1 on Days 12, 13, and 15; 2 on Day 16; 3 on Day 17; and 4 on each subsequent experimental day. When more than 1 trial per day was run, the intertrial interval for individual *S*s was 10 min.

---

[1] This study was supported in part by Medical Research Council Grant G/967/271B to the senior author.

[2] Requests for reprints should be sent to K. T. Strongman, Department of Psychology, University of Exeter, Gandy Street, Exeter, England.

The test period (Days 41–50) differed from the training period for E(0) and E(2) on half of the 40 possible trials. Thus on 20 of the test trials, E(0) experienced nonreward in GB_1, whereas E(2) encountered an increased reward (two pellets) in GB_1. The experimental trials (T) for E(0) and E(2) were interspersed with normal trials (N) according to the following schedule: NTNT, TNTN, NTTN, TNNT, NNTT, TTNN, TNTN, NTNT, NTTN, TNNT. Thus in each block of 8 trials, half were N and half T, one of each occurring in each of the four daily positions to counteract any possible order effects. During this test period, C(0) and C(2) continued to receive the same number of pellets as previously. On all trials, running time in A_2 was recorded.

Results.—For each block of eight trials (the first block covering Days 12–17, subsequent blocks representing pairs of experimental days), the median running time in A_2 was calculated for each *S*. In the case of E(0) and E(2) during the test period, separate medians (each of four trials within the block) were computed for normal (N) trials and test (T) trials.

Table 1 shows mean median running time in A_2 for E(0) and E(2) during the test period. The normal FE is evident in E(0), running time on nonrewarded trials being significantly faster than on rewarded trials, $F\ (1,\ 16) = 12.51$, $p < .005$. Comparison of the performance of E(0) with that of C(0) revealed no evidence of response suppression, the running times of the two groups not differing significantly on rewarded trials during the test period, $F\ (1,\ 35) = .15$, $p > .05$. Further evidence that, with the quantities of reward used, response suppression due to prefeeding in GB_1 does not occur derives from the lack of significant difference between the test period performance of C(0) and C(2), $F\ (1,\ 35) = 1.16$, $p > .05$. Of greater interest is the performance of E(2), where *S*s received an increased reward in GB_1 on half of their test trials. After the initial block in this test period, *S*s ran faster following their normal reward in GB_1 than after an increased reward. This difference, including the initial reversal, is significant, $F\ (1,\ 16) = 4.49$, $p = .05$. No difference was apparent between test trial performance and the previous training level, implying that a few trials of experience with a greater reward than that given during training are sufficient to provide a "frustrative reward" effect when *S*s are presented with the training reward. Comparison with control group data confirms this, ruling out the possibility of response suppression on higher reward trials. There were no significant differences between E(2) performance on these trials and that of C(2) and C(0) during the test period, $F\ (1,\ 40) = 3.71$, $p > .05$; and $F\ (1,\ 35) = .89$, $p > .05$.

TABLE 1

MEAN MEDIAN RUNNING TIME (IN MSEC.) IN TEST PERIOD FOR E(0) AND E(2)

Trial blocks	E(0)		E(2)	
	Normal	Experimental	Normal	Experimental
11	939	895	1,095	1,033
12	1,003	880	948	1,092
13	925	884	946	1,056
14	873	837	965	974
15	1,028	896	933	992

Discussion.—The present results show that brief experience with increased reward renders frustrative a reward of which *S* has had considerable experience. This is in accord with Bolles (1967), who suggests that incentive effects seem more dependent on *S*s' recent history of reinforcement than on the whole history, but at variance with a study by Roussel (1952). The latter showed that when reward was available in GB_1 on 50% of the trials from the beginning of training, the effects of nonreward were not discernible until after about 20 trials. This discrepancy may be variability in performance at the start of training, masking any differential effects.

The frustrative reward effect is supported by a second experiment of similar design, but using smaller groups, a 24-hr. intertrial interval, and a brief test period. However, the present procedure is not directly comparable with the normal frustration procedure—nonreward in GB_1. The paradigm is more comparable with a less than total reduction of reward to produce frustration. This typically leads to a less pronounced FE than does nonreward (e.g., McHose & Ludvigson, 1965). The frustrative reward effect should be more marked if *S*s were trained with no reward in GB_1 and subsequently rewarded there.

REFERENCES

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *Journal of Experimental Psychology*, 1952, **43**, 363–368.

BOLLES, R. C. *Theory of motivation.* New York: Harper & Row, 1967.

McHOSE, J. H., & LUDVIGSON, H. W. Role of reward magnitude and incomplete reduction of reward magnitude in the frustration effect. *Journal of Experimental Psychology*, 1965, **70**, 490–495.

ROUSSEL, J. Frustration effects as a function of repeated nonreinforcements and as a function of the consistency of reinforcement prior to the introduction of nonreinforcement. Unpublished study, 1952. Cited by A. Amsel, The role of frustrative nonreward in non-continuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

WAGNER, A. R. Role of reinforcement and nonreinforcement in an "apparent frustration effect." *Journal of Experimental Psychology*, 1959, **57**, 130–136.

(Received February 8, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 185–187

CORRECTION OF FALSE MOVES IN PURSUIT TRACKING

RONALD W. ANGEL[1] AND JOSEPH R. HIGGINS[2]
School of Medicine, Stanford University

Ten *S*s were tested on a pursuit tracking test under display conditions that caused them to make false moves. The mean latency of the false moves was longer than the mean latency of correct moves. The mean time needed to arrest false moves was less than the mean time needed to initiate any motor response, whether correct or incorrect. Forty-one percent of the false moves were arrested in less than the estimated minimum time needed to process visual feedback. The data confirm previous reports that *S*s are able to correct errors more quickly than they can respond to external stimuli.

Several experiments have shown that when two stimuli are separated by an interval of .5 sec. or less, the latency of the second response is longer than that of the first. In order to explain this finding, Welford (1952) has suggested that sensory feedback from the first response may "capture" the central mechanisms for a brief period and thereby delay the processing of a second stimulus. More recently, however, Gibbs (1965) has found that incorrect motor responses can be amended within .1 sec. from the time they begin. This observation shows that feedback from one movement does not necessarily increase the latency of a succeeding response. On the contrary, Gibbs inferred that the rapid corrective responses were facilitated by proprioceptive feedback from the incorrect moves. The experiments reported in the present paper confirm the work of Gibbs.

Method.—Five men and five women, of ages ranging from 18 to 45, used a joy stick to position a cursor line on an oscilloscope and track a target line. Hand movement to the right or left was detected by means of a potentiometer and a tachometer, which were coupled to the joy stick. Voltages proportional to position and velocity of the hand were recorded on magnetic tape along with a signal showing position of the target line. The tape-recorded signals were displayed by means of a strip-chart recorder, and measurements were made on the paper (see Fig. 1 and 2).

The position of the target line was controlled by *E*, who operated a hand switch, causing the target to occupy one of three positions: the center of the display tube, 3 cm. to the right, or 3 cm. to the left. The track consisted of 140 target steps, which appeared random but actually followed the same order for every *S*. When the target was centered, it jumped to the right or left with equal probability, and each lateral jump was followed by a return to center. Every block of 20 steps contained 5 jumps to the right, 5 to the left, and 10 back to center.

A second switch, also operated by *E*, determined the directional relation between control and display. When this switch was in the "positive" position, the joy stick and the display cursor moved in the same direction. When the switch was in the "negative" position, the joy stick and the display cursor moved in opposite directions. On the first 40 target steps, the control–display relation was positive. On Steps 41–80, the relation was negative. For Steps 81–120, the relation was changed after each block of 10 moves, as illustrated in Fig. 3. For the final 20 steps, the control–display relation was changed after every recentering move.

The *E*, who was seated behind *S*, monitored the display tube and made sure that the display cursor was superimposed on the target before switching the target line to a new position. Since the different *S*s required variable times to complete their moves, it was not feasible to standardize the interval between the completion of one response and the time of the next target step. In view of the fact that temporal uncertainty affects reaction time, the intervals between successive target steps were recorded

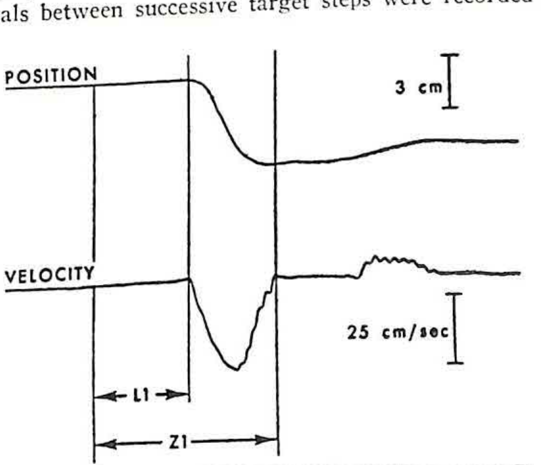


FIG. 1. Tracking response with incompatible control–display relation. (Top line: position of joy stick. Second line: velocity. Downward deflections indicate move to right. As trace begins, velocity is zero. At the time shown by step on lower line, target jumps 3 cm. toward the left. After a latency of 200 msec. [L1], joy stick begins to move to the right. Velocity returns to zero level 380 msec. after target step [Z1]. Duration of the initial move is 180 msec. [Z1 minus L1].)

---

[1] Also at Veterans Administration Hospital, Neurophysiology Laboratory.

Requests for reprints should be sent to Ronald W. Angel, Veterans Administration Hospital, 3801 Junipero Serra Boulevard, Palo Alto, California 94304.

[2] The authors wish to thank Byron W. Brown, Jr., for help with the statistical analysis and Norman H. Mackworth for helpful su[illegible]s.

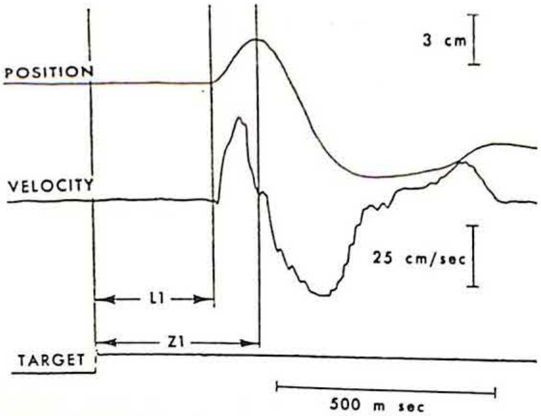


FIG. 2. Incorrect tracking response. (Step on bottom line shows time when target jumps 3 cm. to left. After a latent period of 267 msec. [L1], joy stick begins to move leftward and the response marker [not shown] moves rightward, adding to the error generated by track. The false move is arrested at time Z1, 367 msec. after target step. Duration of false move is 100 msec. [Z1 minus L1].)

and measured along with the other data. The interstep intervals were found to range between 2 and 4 sec., but the variance was approximately the same for different *S*s and different parts of the track. Hence none of the effects reported can be ascribed to differences of temporal uncertainty.

Results and discussion.—Each of the 10 *S*s made 140 moves, so that 1,400 moves were recorded altogether. Of this total, 141 were found to be false moves (FMs), defined as moves in which the initial acceleration caused the response marker to move away from the target. The number of FMs varied from 6 to 21 per *S*, the group mean of 14.1 being very close to 10% of the total series. The errors were not distributed randomly throughout the track, but tended to occur when the polarity of the control-display relation was negative (see Fig. 3). This agrees with Gibbs' (1965) finding that the frequency of errors was much greater when the incompatible display was used.

In order to determine the mean reaction times (RTs) for correct moves (CMs), measurements were taken from all correct responses made on Steps 71–80, which were selected to represent optimal performance with the negative display polarity. The mean RTs ranged from 237 to 364 msec., the group mean being 300 msec. The mean RTs for FMs ranged from 245 to 480 msec., the group mean being 357 msec. The difference of the group means for CMs and FMs was significant, $t\ (9) = 3.2$, $p < .01$. This result agrees with Gibbs' (1965) finding that on probable and unequivocal steps, the errors were associated with longer latencies, i.e., the opposite of the expected relation between accuracy and speed.

The duration of any move was defined as the time from onset of movement to the time when velocity became zero (see Fig. 1 and 2). The mean

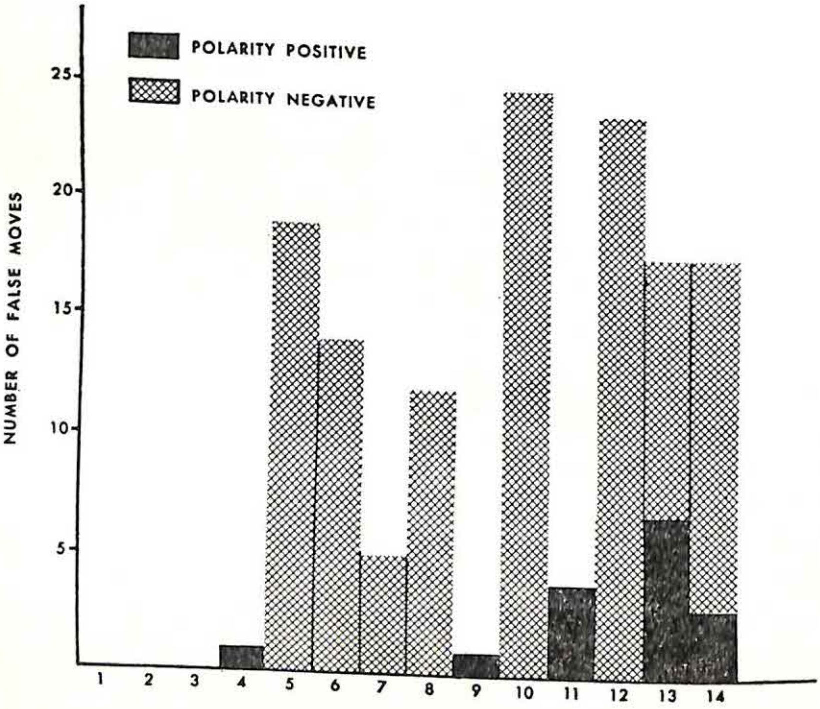


FIG. 3. Numbers of false moves occurring in different portions of the tracking task. (The probability of obtaining a false move depends on the polarity of the feedback display. The moves in Blocks 1–4, 9, and 11 were made under normal display conditions, and relatively few false moves occurred. The moves in Blocks 5–8, 10, and 12 were made with the incompatible display, and many false moves were obtained. In Blocks 13 and 14, the polarity of the display was changed after each pair of target steps.)

duration of the CMs in response to Target Steps 71–80 varied from 168 to 571 msec., the group mean being 362 msec. The mean duration of FMs varied from 156 to 276 msec., the group mean being 215 msec. This is 147 msec. less than the mean for CMs, $t\ (9) = 4.5$, $p < .005$.

If *S*s can amend their own FMs within so short a time, are the corrective responses based on visual feedback from the display? Keele and Posner (1968) have found that the minimum amount of time necessary to process visual feedback from a movement is 190–260 msec. In the present experiments, the group mean for amendment times was 215 msec., but for three *S*s the mean amendment times were less than 190 msec. Moreover, 41% of all FMs made by the 10 *S*s were arrested within less than 190 msec.

If some FMs are arrested before any visual feedback could become effective, what provides the stimulus for the very fast corrective responses? Gibbs (1965) postulated that *S*s use proprioceptive monitoring to provide the fast correction times. An alternative possibility is that the brain monitors and records outgoing motor impulses, amending those which do not agree with the intended plan of action.

The findings confirm the work of Gibbs (1965) and also agree with Rabbitt's (1966) conclusion that error correcting responses are faster than equivalent, correct responses. It appears that *S*s' internal monitoring of their own responses may allow them to correct errors more quickly than they respond to any external signal from the display. No evidence was found to support the view that proprioceptive feedback delays the response to stimuli arriving during movement.

REFERENCES

GIBBS, C. B. Probability learning in step-input tracking. *British Journal of Psychology*, 1965, **56**, 233–242.

KEELE, S. W., & POSNER, M. J. Processing of visual feedback in rapid movements. *Journal of Experimental Psychology*, 1968, **77**, 155–158.

RABBITT, P. M. A. Error correction time without external signals. *Nature*, 1966, **212**, 488.

WELFORD, A. T. The psychological refractory period and the timing of high-speed performance—a review and a theory. *British Journal of Psychology*, 1952, **43**, 2–19.

(Received February 10, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 187–189

EFFECTS THAT A MASSED REPETITION OF ONE PAIR HAS ON OTHER PAIRS IN A LIST[1]

GEORGE W. McCONKIE[2]

Department of Education, Cornell University

It was hypothesized that a massed repetition of one word pair in a list facilitates the recall of other pairs in that list. This was confirmed using six-pair lists with Murdock-type tests of individual pairs in the lists. A massed repetition at Serial Positions 1, 3, or 4 resulted in better recall of the list as a whole, and there was a striking absence of peaks in the recall curves at points corresponding to the repeated pairs.

In the study of verbal learning, repeating information generally increases the likelihood that *S* will recall it. Recently, a limitation to this law has been reported in studies which have found that when two presentations of the same item follow each other in immediate succession, the effect of the repetition may be greatly reduced or even eliminated entirely (Greeno, 1964). Theories proposed to account for this effect (Atkinson & Shiffrin, 1968; Landauer & Rubin, 1966) have assumed that less total learning occurs when repetitions are massed. An alternative hypothesis is that the massing of presentations of the same item results in changes in the rehearsal pattern for the list, so the extra time comes to benefit not the repeated item, but other members of the list. Specifically, it might be expected that *S*s spend their time during the massed repetition of one pair in rehearsing previously presented pairs, thus strengthening them rather than the repeated pair.

The present study was designed to test whether a massed repetition of one word pair in a list raises the recall probabilities of other pairs in that list, as suggested previously. This was accomplished by presenting six-pair lists with a massed repetition of one of the pairs in some lists. Recall for all items in the list was tested with the method used by Murdock (1964).

Method.—Sixty lists of six word pairs each were constructed by randomly pairing common English words (frequency between 50 and 100 per million according to the Thorndike-Lorge, 1944, word count) and arbitrarily assigning these pairs to lists in such a way that no two stimulus words in a list began with the same letter.

---

[1] The data in this experiment were reported in a paper presented at the meeting of the Eastern Psychological Association, Washington, D. C., April 1968.

[2] Requests for reprints should be sent to George W. McConkie, Department of Education, Cornell University, Ithaca, New York 14850.

The lists were projected on a screen one pair at a time. There were five presentation conditions: either the first, third, fourth, sixth, or none of the pairs in the list was presented twice in immediate succession (Cond. D1, D3, D4, D6, and 6, respectively). Five groups of *S*s each received a different order of these presentation conditions, but with the lists of pairs always in the same order. Thus, each list was presented equally often under each presentation condition. Following the presentation of a list, each *S* drew from an envelope a slip of paper containing a single word. He then attempted to recall the word which was paired with that stimulus and he wrote the response on the paper. Since each list had six pairs, one-sixth of the *S*s were tested on each of the pairs following the presentation of each list. This was assured by preparing six recall orders, among which *S*s were equally divided. The five presentation conditions and six recall conditions yielded a total of 30 conditions. Each *S* was tested 60 times, twice under each condition. A total of 90 *S*s, all Cornell University undergraduates, were tested in groups of from 5 to 10 each.

The word pairs were presented for 2 sec. each, with 1 sec. between pairs. Nine seconds were allowed for recall, and there was a 9-sec. rest prior to the beginning of the next list.

Results.—The data are summarized in Fig. 1, which shows the proportion of the pairs recalled at each serial position under each presentation condition. Each point summarizes the data from 180 observations. Arrows indicate the pairs which received massed repetitions. The most striking feature of these curves is the absence of peaks corresponding to repeated items. It is clear that repeating a word pair has a more general effect on the shape of the curve, rather than simply raising the probability of recalling the repeated pair itself. Those lists in which the first, third, or fourth pair was repeated show a general elevation of recall probability for the first four serial positions over the control condition. On the other hand, the control list was actually recalled best at Serial Positions 5 and 6. Repeating the final pair resulted in a general decrement in recall, rather than in a facilitation as found when other pairs were repeated.

A five-way analysis of variance was carried out with two factors between *S*s (order of the presentation conditions and order of the recall conditions) and three factors within *S*s (first 30 vs. second 30 lists, presentation condition, and serial position tested). Two main effects were significant: the serial position effect, $F\ (5,\ 300) = 156.07$, $p < .001$, and the presentation condition effect, $F\ (4,\ 240) =$

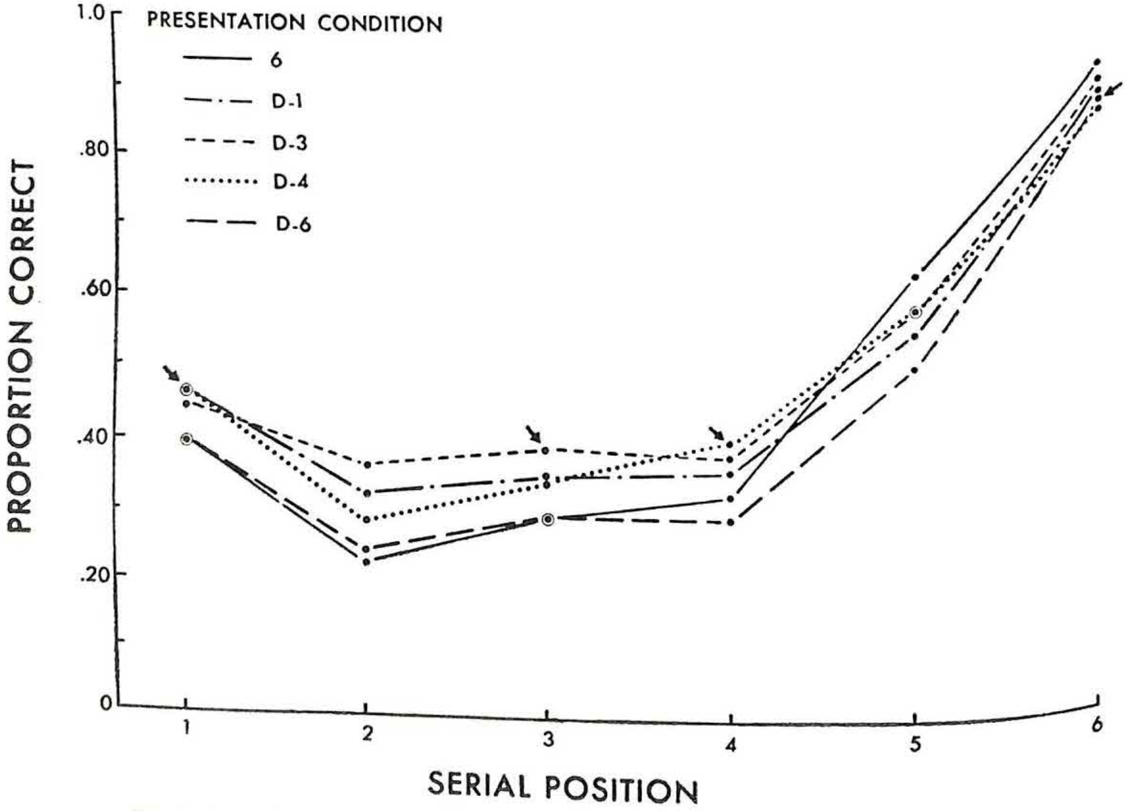


FIG. 1. Proportion of word pairs recalled correctly. (Arrows indicate points corresponding to pairs which received a massed repetition.)

5.59, $p < .001$. A Newman-Keuls test among the presentation conditions yielded the following pattern:

Presentation condition:	D6	6	D1	D3	D4
Proportion correct:	.434	.466	.489	.489	.513

Proportions with a common underline correspond to means which are not significantly different at the .05 level. The Serial Position × Presentation Condition interaction was not significant, F (20, 1,200) $= 1.50$, $p > .05$, indicating again that the effect of a massed repetition is on the entire list, rather than simply on the item repeated.

The Serial Position × First Thirty vs. Second Thirty Lists interaction was significant, resulting mainly from a disappearance of the primacy effect over trials, a commonly observed phenomenon (Murdock, 1964). Other higher order interactions which were significant involved recall order. Since the word pairs were always in the same serial positions during the presentation, *S*s receiving the six different recall orders were being tested on different sets of word pairs, which probably gave rise to these interactions.

Discussion.—The results of the present study indicate that the massed repetition of one pair in a list can facilitate the recall of other pairs in that list. In fact, the effect on other pairs can be as large as that on the repeated pair itself. Also, this effect does not seem to be restricted to just those pairs prior to a repeated pair, as might be predicted by a simple rehearsal theory.

Two aspects of the data are particularly puzzling. It is not obvious why the control condition should produce the best recall at Serial Positions 5 and 6. In view of the absence of a significant Serial Position × Presentation Condition interaction, only future studies can indicate whether this is a stable phenomenon under these conditions. Also, it is difficult to understand why the recall for Serial Position 5 was so poor for Cond. D6. It is clear that whatever *S*s were doing during the repetition of the sixth pair in the list, they were not effectively rehearsing the immediately prior pair.

If further research should indicate that while the massed repetition of a pair results in increments to other pairs in the list, a spaced repetition tends to concentrate the effect on the repeated pair itself, this would serve as a resolution of the discrepancy between the massed vs. distributed repetition effect and the total-time hypothesis (Cooper & Pantle, 1967). About the same amount of learning may occur under both massed and distributed presentations so far as the total list is concerned, but if only the recall for repeated items is measured, the data may show much greater learning for the distributed than for the massed pairs.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968.

Cooper, E. H., & Pantle, A. J. The total-time hypothesis in verbal learning. *Psychological Bulletin*, 1967, **68**, 221–234.

Greeno, J. G. Paired-associate learning with massed and distributed repetitions of items. *Journal of Experimental Psychology*, 1964, **67**, 286–295.

Landauer, T. K., & Rubin, N. Spacing of item repetitions in paired-associate learning. *Proceedings of the 74th Annual Convention of the American Psychological Association*, 1966, **1**, 91–92.

Murdock, B. B. Proactive inhibition in short-term memory. *Journal of Experimental Psychology*, 1964, **68**, 184–189.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received February 13, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 1, 190–192

FURTHER TEST OF THE USE OF IMAGES AS MEDIATORS[1]

GEOFFREY KEPPEL[2] AND BONNIE ZAVORTINK
University of California, Berkeley *Georgetown University*

The *S*s encouraged to use a mnemonic rhyme ("1 is a bun, 2 is a shoe," etc.) were facilitated in their learning of four sets of number-word pairs when compared with *S*s not so instructed. In addition, test-list recall was superior for the mnemonic group. These results generally support those reported by Bugelski, although certain differences between experiments are noted and discussed.

Bugelski, Kidd, and Segman (1967) reported an experiment in which *S*s were instructed to use a prelearned mnemonic rhyme ("1 is a bun, 2 is a shoe," etc.) in making up bizzare visual images in order to recall a series of 10 words. Under the slowest study rate (8 sec. per word), the experimental (E) group, which was instructed to use the mnemonic, achieved nearly perfect recall (97%), while the uninstructed control (C) groups did considerably worse (67%). In a later study, Bugelski (1968) found a continued superiority of the E group on successive word series, although the effect tended to diminish with successive lists and was no longer reliable on the last series.

The present experiment, conducted independently of the latter study, also investigated the persistence of the instructional variable, with some potentially important methodological changes from the earlier studies. With the exception of one comparison that is mentioned later, however, these changes did not affect materially the results reported previously. Thus, they are listed briefly: (*a*) Each list was learned to a criterion of one perfect recitation instead of the single study-recall sequence used by Bugelski (1968). (*b*) Number stimuli were presented during the study periods; it is not clear from Bugelski's descriptions whether or not the testing stimuli were given explicitly in his experiments. (*c*) The importance of the stimulus ordering during pretraining and testing was assessed; Bugelski always presented the material in an ascending order, which was necessary to define the ordinal positions of the words in the series. (*d*) For the first test list, the C group included here approximated an A-B, C-D control for nonspecific transfer. In comparison with his E group, Bugelski's two C groups (cf. Bugelski et al., 1967) may have suffered in one case from negative transfer and in the other from an absence of nonspecific transfer.

Method.—There were two C and four E groups, the former being pretrained on 10 pairs of rhyming words and the latter on 10 number-rhyme pairs. All *S*s then received four test lists in which the numbers 1–10 were stimuli and unrelated words were responses. One of the C groups studied the number-word test lists in ascending numerical order, while the other received a random but consistent ordering of the pairs. Two of the E groups received the ascending order of presentation and two the random. In addition, the E groups were further subdivided with regard to the consistent ordering of the number-rhyme pretraining, in ascending order for one and random order for the other.

Pretraining materials for the E groups consisted of the 1-*bun*, 2-*shoe* rhyme listed by Bugelski (1968). The same response words were used in the C group list, but a word that rhymed with the response was substituted for each number, e.g., PUN-*bun*, CLUE-*shoe*, etc. The responses of the test lists were two-syllable concrete nouns of the 1–10 frequency range chosen from the Thorndike-Lorge (1944) count. Four sets of 10 words were selected randomly, and two different pairings of numbers and words were formed. Each pairing was used equally often.

For pretraining in the random presentation conditions of the E groups and in the two C groups, two different presentation orders were used, each for half of the *S*s of a group. Since the numerical ordering of the stimuli on the study trials remained the same on all trials for the ascending groups, the same random ordering was used for the random groups with each list they received. (For these latter groups, however, the particular arrangement was different for each successive test list in order to avoid the establishment of stable serial position cues across the lists.) On the other hand, testing order changed on successive trials for all groups. Four different random test orders were constructed for each list. Four different sequences of the four test lists were used so that lists were equally represented at each stage of practice.

The pairs were presented on a Stowe memory drum and an alternating study-test procedure was employed. The pretraining was carried to three successive perfect trials for all groups, with a 2-sec. rate for both study and test trials. Prior to the administration of the test lists, the E groups were instructed to use the pretraining list in forming

---

[1] This research was supported by Research Grant MH-10249 from the National Institutes of Health, United States Public Health Service, and was conducted at the Institute of Human Learning, which is supported by a grant from the National Institutes of Health.

[2] Requests for reprints should be sent to Geoffrey Keppel, Department of Psychology, University of California, Berkeley, California 94720.

TABLE 1

MEAN NUMBERS RECALLED ON TRIAL 1 AND MEAN TRIALS TO CRITERION ON FOUR TEST LISTS

Comparisons	Test lists			
	1	2	3	4
Trial 1				
Control	4.84	4.34	4.69	4.59
Experimental	6.95	6.52	6.52	6.39
Criterion				
Control	3.38	3.47	3.53	3.66
Experimental	2.66	2.88	2.73	2.78

bizarre images between the stimuli and responses of the test lists; the C groups were instructed merely to learn the pairs. The test lists were learned to a criterion of one perfect recitation, with a 6-sec. presentation rate and a 4-sec. test rate. Approximately 2 min. elapsed between successive lists. Following the learning of the last test list, *S*s were given 5 min. to write down the words of the test lists opposite the appropriate stimulus members.

There were 16 *S*s per group, half of whom were students fulfilling class requirements and half of whom were paid volunteers obtained through the student placement bureau. The six conditions were randomized separately 16 times; list orders and pairings were balanced over the 16 running blocks. The *S*s were assigned to a particular condition in order of arrival for the experiment.

Results and discussion.—For the cases in which the ordering of the stimulus materials did not affect performance, the data presented ignores these classifications, although they are still retained in the statistical analyses.

The two pretraining tasks proved to be of comparable difficulty, with the combined E groups requiring 5.39 trials and the C groups 5.94 trials, F (1, 90) = 1.12.

First-trial performance and criterion scores on successive lists for the combined E and C groups are listed in Table 1. With both measures there was a clear superiority of the E condition over the four lists combined, $F = 36.50$ for Trial 1 performance and 13.29 for criterion performance. Contrary to the findings of Bugelski (1968), however, there was no suggestion of a convergence of the two groups on the later lists. It is possible that the pronounced improvement shown by the C group in his experiment may have been due to the fact that this group received no pretraining, so that practice effects during test-list learning may have exceeded interference effects. In addition, the C groups did not show the saw-tooth function that was observed by Bugelski. This discrepancy between experiments may be attributable to use of a criterion on the test lists in the present study, a procedure which would tend to more nearly equate degree of interlist interference than would Bugelski's single study-test trials.

The results of the recall test, scored in terms of correctly paired responses, are shown in Fig. 1.

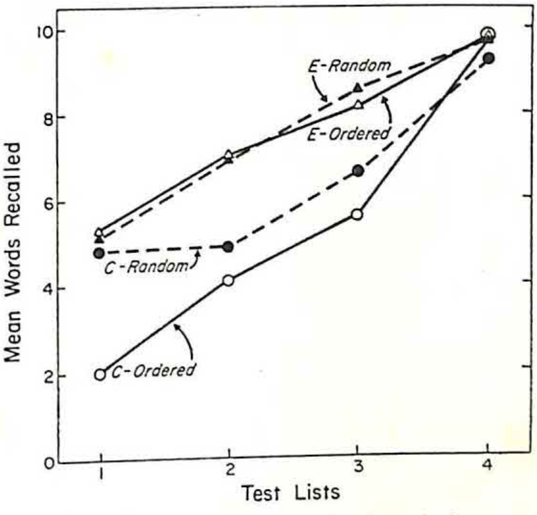


FIG. 1. Mean number of words recalled from the four sets of test lists.

As would be expected, recall increased as a function of list position, $F = 171.80$. More importantly, the E groups gave significantly more pairs than did the C groups, $F = 29.38$, and there were two significant interactions both involving list position. More specifically, a significant two-way interaction, F (3, 270) = 7.95, $p < .01$, obtained by collapsing across presentation order, is the result of equivalent recall of the last list by the E and C groups and superior recall by the E group on the earlier lists. This conclusion must be tempered somewhat because of the significant three-way interaction of list position, the E/C variable, and stimulus ordering of the test lists, F (3, 270 = 4.08, $p < .01$. Essentially, the order of the stimuli on the test lists had no effect for the E groups, but was a relevant variable for the C groups, with higher recall being found on the early lists by the group receiving a new random ordering of the stimuli on successive lists.

These results deviate from those reported by Bugelski (1968) in several important ways: (*a*) There is no difference between the present E and C groups on the final list, but a reliable one for Bugelski, a discrepancy which is undoubtedly due to the fact that *S*s were taken to a common performance criterion in the present study. (*b*) Both the E and C groups show increasingly poorer recall of the earlier lists; Bugelski found only small differences among the lists for the E group, but a sizable drop for the C group. (*c*) The E-C difference remained relatively constant over the earlier lists when the ordered groups are considered, but was nearly eliminated in first-list recall when the two random-presentation groups are compared; while not evaluated by Bugelski, the E-C difference he obtained was larger for the earlier lists.

For recall of the first list, a comparison of the E and C groups receiving the more typical, i.e., random, stimulus presentation revealed equivalent performance. In this case, then, it appears that associations formed by the E group are not necessarily more resistant to retroactive inhibition (RI) than those formed by the C group, as has been suggested by Bugelski (1968) in his interpretation of his data. In contrast, however, a large difference is present for the two groups receiving the consistent order of presentation on successive lists, the procedure employed by Bugelski. It is possible that *S*s in the consistent C group tended to use serial position cues in learning to a greater extent than did *S*s in the consistent E group and that the constancy of this position cue on successive lists introduced an additional source of interference for this particular group (cf. Keppel, 1964).

In general, greater RI had been expected for the *E group* because of the fact that "forcing" *S*s to use the same coding cues in the free recall learning of two lists increases the amount of RI that is observed (cf. Zavortink & Keppel, 1968). Presumably, then, if *S*s in the E group are effectively forced to use the same image or mediator on successive lists, a similar increase in RI should be found. In retrospect, of course, it really is not known for which group (E or C) the successive mediators are more similar in view of the fact that Bugelski's (1968) *S*s reported the use of fairly different mediators for each peg word over the six lists. As it now stands, therefore, no conclusion can be drawn from these data in this regard until more is known concerning the functional similarity of the interlist mediators for the two groups.

REFERENCES

Bugelski, B. R. Images as mediators in one-trial paired-associate learning. II: Self-timing in successive lists. *Journal of Experimental Psychology*, 1968, **77**, 328–334.

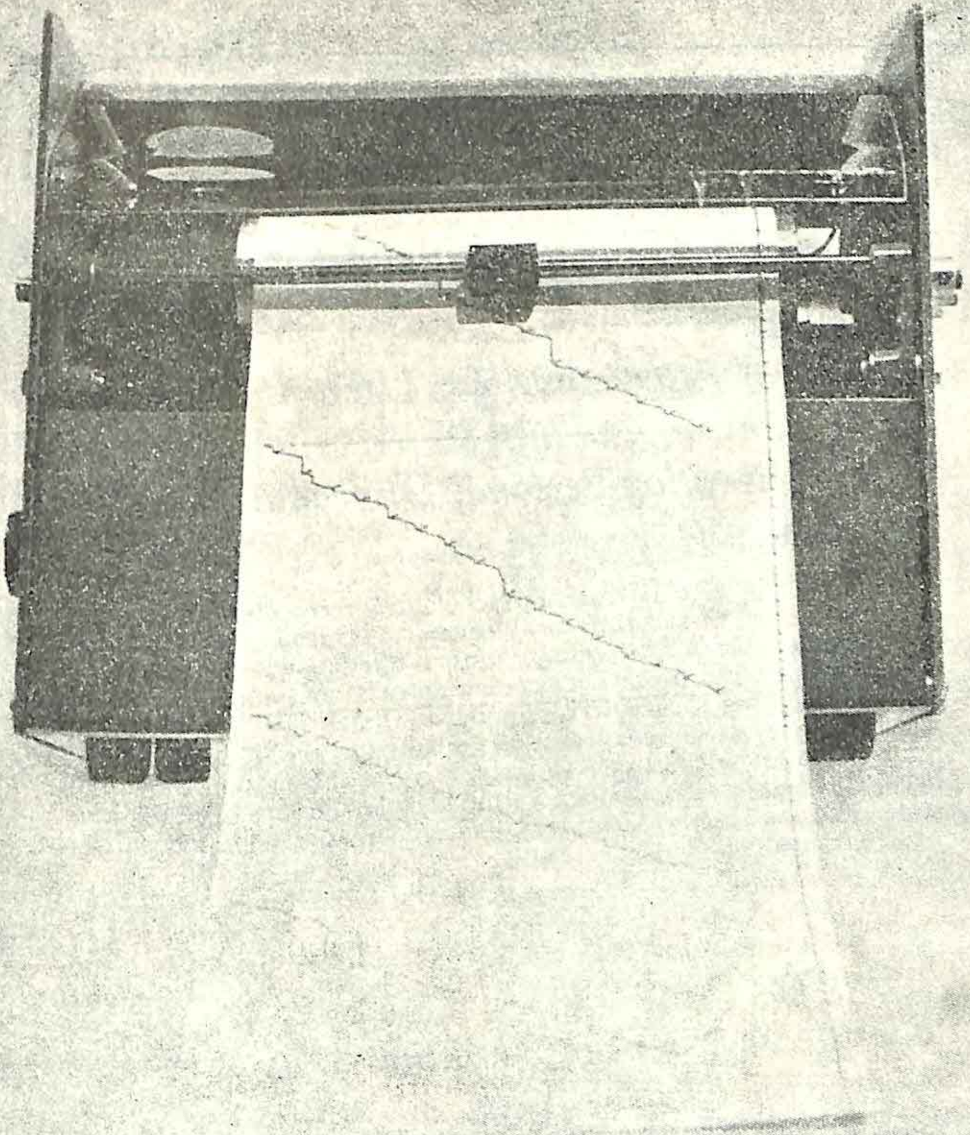
Bugelski, B. R., Kidd, E., & Segmen, J. The image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology*, 1967, **76**, 69–73.

Keppel, G. Retroactive inhibition of serial lists as a function of the presence or absence of positional cues. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 511–517.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

Zavortink, B., & Keppel, G. Retroactive inhibition in free-recall learning with alphabetical cues. *Journal of Experimental Psychology*, 1968, **78**, 617–624.

(Received February 17, 1969)



RELIABILITY

MAKES THE DIFFERENCE

Models C-3, C-3SHS
Harvard Cumulative Recorders

Ralph Gerbrands Company
8 BECK ROAD ARLINGTON, MASSACHUSETTS 02174 (617) 648-6415

New

Volume I
The Measurement of Psychological States Through the Content Analysis of Verbal Behavior

Volume II
Manual of Instructions for Using the Gottschalk-Gleser Content Analysis Scales: Anxiety, Hostility, Social Alienation-Personal Disorganization

Louis A. Gottschalk *and* **Goldine C. Gleser**

A major need in psychiatric and psychological research has been the development of methods for the accurate measurement of transient emotional, motivational, and cognitive states. The present volumes offer such a method based on language as the raw material from which measurements can be made. Drawing on psychoanalysis, learning theory, and linguistics, the authors present a series of scales, which include categories of crucial verbal expressions, by means of which nonprofessional technicians can objectively score the content of verbal behavior as it is expressed in a five-minute verbal sample. The magnitude of various mental states, such as anxiety, hostility, cognitive and intellectual impairment, can thus be determined.

1969 LC: 69-15828 Vol. I 317 pages $12.50
Vol. II 176 pages $6.00

The Measurement of Sensation

A Critique of Perceptual Psychophysics

C. Wade Savage

The logic of measurement in psychophysics has been a sore point in psychology—and in the philosophy of science—since the middle of the nineteenth century. Here, in an acute study of the distinctions between direct and indirect measurements, Mr. Savage breaks away from the main tradition of the philosophy of perception and formulates an original approach to the measurement of psychophysical sensations, in particular the sensations of loudness and pitch, afterimages, and pain.

1969 LC: 69-15941 260 pages $15.00

from California

University of California Press • Berkeley 94720

We'll design a digital logic system for any experimental parameters you think up.

No cost. No obligation.

From the time BRS-Foringer first introduced solid-state logic modules for life science experiments, personalized system design assistance has been our hallmark.

All BRS-Foringer sales personnel have solid backgrounds in psychology, medicine, engineering or education. They talk your language. Understand your requirements. And they are especially experienced in designing digital logic systems to fulfill your experimental designs . . . to help you achieve greater accuracy and control in experiments and reduce set-up and maintenance time.

A telephone call or a letter outlining your proposed experiment is all it takes to put us to work.

You'll receive a comprehensive DigiBits system diagram showing how all modules interconnect and operate. And along with the diagram comes a complete list of required equipment and exactly how much it will cost.

Then you decide.

The service is provided without cost; without obligation. We're sure if you select a solid-state system for your experiments, you'll want the best. That's where BRS-Foringer comes back into the picture — with equipment guaranteed for an unlimited period of time plus continuing technical advice.

Sound interesting? Call (301) 474-2900

BRS FORINGER®

A DIVISION OF TECH SERV, INC.

Main Office: 5451 Holland Drive, Beltsville, Maryland 20705
(301) 474-2900

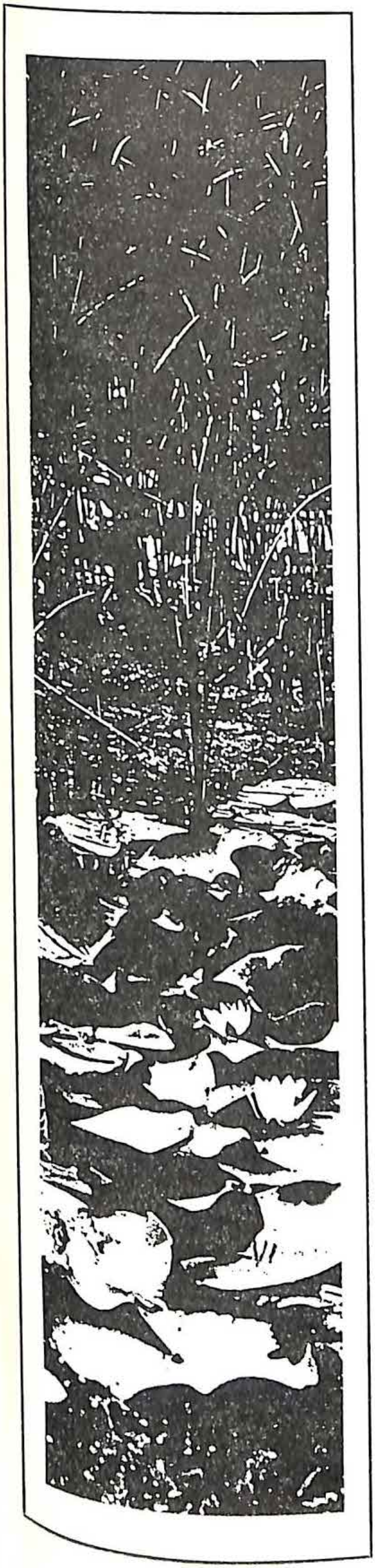
RUSSIAN–TO–ENGLISH

PSYCHOLOGY TRANSLATORS WANTED

You can keep in touch with the latest Soviet research in your field, while supplementing your income, by translating in your own home on a part-time basis. If you have a native command of English, a good knowledge of Russian, and experience and academic training in the field of psychology, you may be qualified for our program. *Call or write now for additional information: TRANSLATIONS EDITOR*

PLENUM

PUBLISHING CORPORATION
227 West 17 St., New York 10011
(Area Code: 212) 255-0713



Allyn and Bacon texts reflecting recent developments in the field

New **1970 / PSYCHOLINGUISTICS James Deese,** The Johns Hopkins University. A pre-publication reviewer, Professor Robert E. Davidson, University of Wisconsin, states: "My overall impression is that the book is very interesting in its presentation, of good quality and should be a success." This text is an introduction to psycholinguistics with emphasis on modern linguistic theory. The study will include an account of grammatical theory and the universals of grammar; sections on the biology of language, the scaling techniques and other psychological measuring devices applied to the study of meaning; and recent experimental literature on problems in linguistic competence. Paperbound Est. 176 pp.

FUNDAMENTAL EXPERIMENTAL DESIGN Jerome L. Myers, University of Massachusetts. Thoroughly class-tested, the material combines an emphasis on understanding basic principles of design and analysis with detailed development of computational formulas and an extensive presentation of numerical examples. The text requires no mathematics beyond college algebra and begins with a review of basic principles of experimentation and statistical inference. 6 x 9 416 pp.

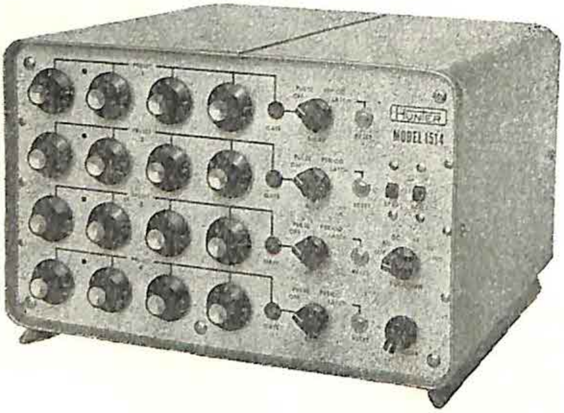
GENERAL PSYCHOLOGY James Deese, The Johns Hopkins University. This highly acclaimed beginning psychology text provides the introductory student a readable and stimulating overview of general psychology. 7⅜ x 10½ 649 pp.

READINGS IN GENERAL PSYCHOLOGY Wayne H. Bartz, Iowa State University. The Bartz book of readings is indexed to Deese' Psychology, but can be used effectively with any general psychology text. The articles exemplify contemporary methods of investigation and theoretical orientations in psychology. Paperbound 680 pp.

THE CAUSES OF BEHAVIOR: Readings in Child Development and Educational Psychology, Second Edition Judy F. Rosenblith, Wheaton College; and **Wesley Allinsmith,** University of Cincinnati. The highly successful first edition has been revised and now includes recent research in child development and educational psychology, a broader discussion of culturally deprived children and a thorough treatment of Piaget. Designed for courses that combine human development and educational psychology. 6 x 9 592 pp.

470 Atlantic Avenue, Boston, Mass. 02210

HUNTER INSTRUMENTS:
THE RESULT OF TWENTY YEARS EXPERIENCE

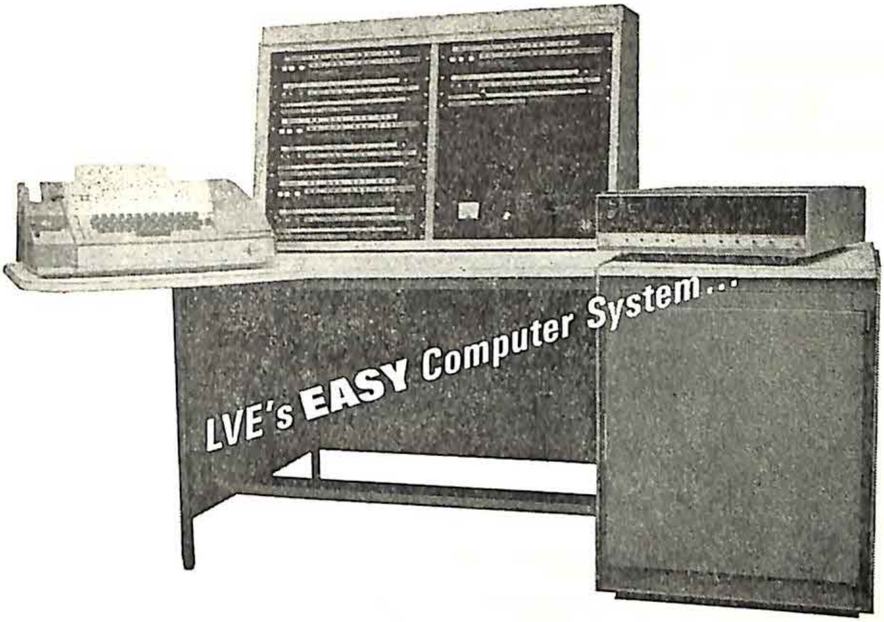


FULL LINE OF TIMING AND COUNTING EQUIPMENT

- Inexpensive Vacuum Tube Models
- Advanced Solid-State Models
- Multi-Channel Instruments
- Photo-Relay Equipment
- Animal Runway Modules
- WRITE FOR OUR FREE CATALOG

Hunter Manufacturing Company, Inc.

P.O. Box 6566, Iowa City, Iowa 52240



interAct

COMPUTER INTERFACE

INTERACT, an INTERface system between a digital computer and the new ACT language, is a problem-oriented computer system designed to control and data record the sequence of events in Psychological experiments.

It is a completely self-contained system in that no special modifications, extra logic circuitry, or peripheral equipment (other than the standard ASR-33 Teletypewriter) is required for operation.

It is an open-ended system in that it can employ virtually any 4096 word, 12 or longer bit core memory general purpose lab computer.

And finally but probably most important, it is a simple system to use. The language employed, ACT for Automated Contingency Translator was written specifically for Behavioral Scientists. The language is sufficiently simple so that an individual with no computer skills can, in a few minutes, write a program in natural English to produce the most complex procedures. Virtually any set of contingencies can be written and, if they can be stated logically . . . They can be written in ACT.

INTERACT we feel, is the most advanced approach made to date for mating the precise world of the computing art to the ever-growing world of the Behavioral Sciences.

Literature
- ACT programming manual
- INTERACT brochure

The above literature is free upon request.

LEHIGH VALLEY ELECTRONICS, INC.

Box 125 Fogelsville, Pa 18051
Phone (215) 285-4211

441 Waverley Street Palo Alto, Calif 94301
Phone (415) 321-4212

Titles in Experimental Psychology

Personality: A Behavioral Analysis
By Robert W. Lundin, The University of the South

This book is a revision of the author's *Personality: An Experimental Approach.* It contains a significant amount of new material not available when the original volume was published in 1961. The title change reflects the addition of much contemporary research in the application of operant conditioning techniques to human behavior. Such topics as stuttering, human concept formation, autistic children, implicit response, and punishment are discussed. In addition, an entirely new chapter on behavior therapy has been included.

1969, 464 pages, $7.95

Learning: Processes
Edited by Melvin H. Marx, University of Missouri, Columbia

Up-to-date and intensive treatment of each topic is offered by: Melvin H. Marx: "Learning Processes"; M. R. D'Amato, Rutgers University: "Instrumental Conditioning"; I. Gormezano, University of Iowa, and John W. Moore, University of Massachusetts: "Classical Conditioning"; Harry I. Kalish, State University of New York, Stony Brook: "Stimulus Generalization"; Charles N. Cofer, Pennsylvania State University: "Verbal Learning"; Harry C. Ellis, University of New Mexico: "Transfer and Retention"; and Jack A. Adams, University of Illinois, Urbana: "Motor Behavior."

1969, 515 pages, $9.95

Learning: Interactions
Edited by Melvin H. Marx, University of Missouri, Columbia

The following contributors explore the interrelationships between learning and other behavioral functions: Robert C. Bolles, University of Washington: "Interactions with Motivation"; James Vanderplas, Washington University: "Perception and Learning"; Robert L. Dominowski, University of Illinois: "Problem Solving and Concept Attainment"; Edward Zigler, Yale University: "Learning, Development and Social Class in the Socialization Process"; Kenneth H. Brookshire, Franklin and Marshall College: "Comparative Psychology of Learning"; and Shepard Siegel, McMaster University: "The Physiology of Conditioning."

1970, approx. 416 pages, prob. $8.95

Experimental Social Psychology
By Judson Mills, University of Missouri, Columbia

Significant experiments, most of them carried out within the last ten years, are discussed by: Dana Bramel, the State University of New York, Stony Brook: "Interpersonal Attraction, Hostility and Perception"; Ellen Berscheid, University of Minnesota, and Elaine Walster, University of Wisconsin: "Attitude Change"; Charles Kiesler, Yale University: "Group Pressure and Conformity"; Eugene Burnstein, University of Michigan: "Interdependence in Groups"; and Judson Mills, University of Missouri: "The Experimental Method".

1969, 466 pages, $9.95

Write to the Faculty Service Desk for examination copies.

THE MACMILLAN COMPANY, 866 Third Avenue, New York, New York 10022

In Canada, write to Collier-Macmillan Canada, Ltd., 1125B Leslie Street, Don Mills, Ontario

341

25 · 8 · 71 Lie

Journal of Experimental Psychology Monograph

Vol. 82, No. 1, Part 2 October 1969

CONDITIONS OF RECOVERY AFTER UNLEARNING[1]

LEO POSTMAN,[2] KARAN STARK, AND DIANE HENSCHEL

University of California, Berkeley

The conditions and characteristics of the recovery of first-list associations following the end of interpolated learning (IL) were examined in five experiments on retroactive inhibition. In Exp. I–IV, recovery was studied as a function of the paradigm of transfer and the temporal arrangement of original learning (OL), IL, and the test of recall (MMFR). In the experimental treatments, *S*s learned three successive paired-associate lists conforming to either the A-B, A-C, A-D or the A-B, C-D, E-F paradigm. In one experimental arrangement the IL-test interval was varied, and in the other the temporal point of interpolation (TPI) was varied. Both manipulations yielded evidence of first-list recovery: first-list recall increased as a function of the IL-test interval and was greater when IL occurred early in the retention interval than when it immediately preceded the retention test. Recovery was found to depend critically on an adequate degree of OL. The increases in first-list recall were comparable for the two paradigms. Greater changes in performance were produced by manipulation of the TPI than of the IL-test interval. In Exp. V, recovery of backward associations was investigated under the A-B, C-B and A-B, B-C paradigms, and a trend toward recovery was found only for the latter. The results are viewed as consistent with the hypothesis that recovery reflects primarily the dissipation of response set interference, i.e., of the tendency established during IL to suppress the responses from earlier lists.

The question of whether unlearning is a reversible process has critical implications for current analyses of the mechanisms of interference in forgetting. The reduced availability of first-list associations after interpolated learning (IL), which is taken as evidence of unlearning, has been widely attributed to the operation of a process having the functional properties of experimental extinction (cf. Keppel, 1968; Postman, 1961). It follows that unlearned associations, like extinguished conditioned responses, should show spontaneous recovery as the interval between the end of IL and the test of first-list retention is lengthened. Experimental tests of this hypothesis have focused on temporal changes in retention under the A-B, A-C paradigm as measured by the method of modified free recall (MMFR). The MMFR test has been viewed as providing a relatively pure measure of unlearning on the assumption that performance under this procedure is immune to the effects of response competition and cannot be influenced by failures of list differentiation.

The evidence for spontaneous recovery obtained in studies of temporal changes in MMFR performance has been equivocal. For retention intervals ranging from several hours to a week or more, significant rises in first-list recall have typically failed to mate-

---

[1] This research was supported by Grant MH 12006 from the National Institute of Mental Health. The work was carried out at the Institute of Human Learning.

[2] Requests for reprints should be sent to Leo Postman, Institute of Human Learning, University of California, Berkeley, California 94720.

© 1969 by the American Psychological Association, Inc.

rialize, although there have been occasional exceptions (for a review of recent studies, see Postman, Stark, & Fraser, 1968). In such cases, the possibility remains open, however, that recovery is counteracted by extraexperimental forgetting, i.e., unlearned associations may recover and then decline in strength again prior to the delayed test. On that assumption, absolute rises in first-list recall should be found when the retention interval is sufficiently short to minimize extraexperimental forgetting. In a recent series of experiments (Postman et al., 1968), reliable evidence for first-list recovery was obtained in fact after retention intervals of the order of 20 min. Internal analyses of these results suggested, however, that the rises in first-list recall probably did not reflect gains in associative strength as implied by the analogy to spontaneous recovery, but rather were attributable to the dissipation in time of response set interference.

The proposed interpretation of recovery takes its point of departure from the assumption that the basic condition of unlearning is the operation of a mechanism of response selection during IL. This mechanism is seen as exerting a critical influence on the availability of the entire class of first-list responses rather than on the integrity of specific stimulus–response associations. The argument has been developed in detail elsewhere (Postman & Stark, 1969; Postman et al., 1968). The critical implication for the analysis of the conditions of recovery is that the reduced availability of first-list responses on a test immediately after the end of IL reflects the persistence of *S*'s set to limit himself to the responses from the list practiced last and to suppress the responses from prior lists. The assumption that such response set interference contributes to retroactive inhibition (RI) at recall is in accord with the hypothesis of generalized response competition proposed by Newton and Wickens (1956). The present interpretation holds that set-produced interference plays a major role in bringing about the reduced availability of first-list responses at recall. The selective response set, which is established during IL and entails the suppression of earlier responses, is assumed to be reversible and to dissipate over time. Consequently, first-list responses become more available after a retention interval than immediately following the end of IL.

Several findings converged on the conclusion that the rises in first-list recall observed after relatively short retention intervals reflected the dissipation of response set interference: (*a*) Significant evidence of recovery was obtained only on tests requiring reproduction of the responses, i.e., MMFR and uncued free recall. There were no reliable changes in interference when retention was measured by the method of associative matching. It should be noted that this result was all but predetermined by the fact that a test of associative matching yields only a relatively small amount of RI immediately after the end of IL; thus, there is only a very limited opportunity for improvement over time. The differences among tests of retention imply, of course, that specific stimulus–response associations are highly resistant to interference, so that both the initial level and the subsequent reductions of RI are determined in large measure by the availability of the responses per se (cf. Postman & Stark, 1969). (*b*) Recovery is most likely to be observed under conditions designed to maximize the influence of response set interference. Such conditions include the reexposure of the interfering responses at the time of recall and practice on two interpolated lists rather than one. (*c*) When *S*s are required to recall the interpolated as well as the original lists, there are systematic shifts in the order of output of the responses: the priority in the order of recall of the most recently learned list is less pronounced on a delayed than on an immediate test. (*d*) Performance on MMFR tests is susceptible to proactive as well as to retroactive interference (see also Houston 1967b; Koppenaal, 1963). Classical two-factor theory cannot predict proactive effects in MMFR since the most recently learned list is not subject to unlearning and the test is presumably immune to the effects of competition between different responses to specific stimuli. A possible interpretation is that the simultaneous activation of alternative response systems interferes with the recall of both the original and the interpolated material, especially after the set favoring the most recently learned list has begun to dissipate.

The present paper reports a series of experiments that continue the investigation of the conditions governing the recovery of unlearned associations, with special emphasis on the explanatory usefulness of the concept of response set interference. Two problems receive joint consideration in Exp. I–IV: (*a*) the relationship between first-list recovery and the conditions of transfer during IL and (*b*) the internal consistency of temporal changes in interference observed under different experimental arrangements.

The conditions of transfer during IL are assumed to determine what is unlearned. According to the widely accepted analysis of McGovern (1964), both contextual and specific stimulus–response associations are subject to unlearning whenever the A-B, A-C paradigm applies to these component habits. If the mechanism of unlearning is the same for the contextual and the associative components, and in particular if it has the functional properties of extinction, both classes of associations should recover over time. Thus, the paradigm of transfer may be taken to specify not only what is unlearned, but also what will recover. It follows that the amount of recovery should vary systematically with the conditions of IL. The implications of this analysis come into focus when the A-B, A-C and A-B, C-D paradigms of paired-associate transfer are contrasted. Both specific and contextual associations are assumed to be unlearned under the former, only contextual associations under the latter. The fact that RI on an immediate MMFR test is typically greater for A-C than for C-D is consistent with this interpretation. However, if specific as well as contextual associations recover, differential increases in first-list recall as a function of time should serve to reduce the difference between the two paradigms. That is, the degree of unlearning and subsequent recovery of contextual associations should be equal for the two paradigms; rises in the availability of specific associations should be an additional source of gain in the A-C condition. If first-list recall increases over time but the levels of performance under the A-C and C-D paradigms fail to converge, the validity of the model postulating extinction and recovery of independent classes of associations is called into question. Such a finding would be consistent, however, with the assumption that the degree of suppression of first-list responses is directly related to the amount of associative interference during IL (cf. Postman et al., 1968, p. 690). The greater the degree of suppression, the more slowly response set interference should dissipate. Hence the gain after a retention interval would not be expected to be proportional to the amount of loss on the immediate test, and it would be possible for the difference between the A-C and C-D conditions to remain essentially unchanged even when there is an appreciable amount of first-list recovery.

In the present experiments the amount of first-list recovery for the A-C and C-D paradigms was compared over a retention interval of the order of 20 min. This interval was chosen for the reasons discussed earlier. Interpolated learning consisted of practice on two successive lists conforming to the same paradigm of transfer, a treatment found earlier to be favorable to the detection of temporal changes in interference. In the course of the experiments, recovery was measured after two different degrees of first-list learning. Thus it was possible to examine the relation between the level of immediate recall and the amount of subsequent recovery.

Apart from the comparison between paradigms, evidence was obtained concerning the generality of the phenomenon of recovery. While recovery under the A-C paradigm had been observed earlier, temporal changes in interference under the C-D paradigm remained to be demonstrated. The available findings concerning recovery in the C-D condition are inconclusive. Houston (1967a) failed to observe a change in RI after an interval of 1 wk. However, the level of recall on the delayed test was quite low, with the control group showing a loss of more than 50%. As is generally true of long-term tests, the conditions were clearly unfavorable to the detection of increases in first-list recall. Saltz and Hamilton (1967) likewise found no reliable changes in recall for the C-D paradigm over intervals ranging from 1 min. to 24 hr. However, the initial retention loss was minimal (only slightly more than half an item), so that first-list recall could increase by only a negligible amount. It is interesting to note in passing that a small rise was observed in fact after a 1-hr. interval. In any event, the authors' conclusion from these data that contextual associations do not recover is inappropriate. Essentially the same limitations apply to the results of a study by Abra (1968), who found slightly higher recall for the C-D than the control condition on an immediate test and a small increase in RI after 24 hr. Finally, a methodological question relevant to these studies deserves brief mention at this point. In the experiments of Houston and of Saltz and Hamilton, nonsense syllables were used as the stimuli. In view of the inevitable overlap of letter elements between successive sets of nonsense syllables, there is at least some formal similarity between the stimuli in the two lists, and the defining conditions of the A-B, C-D treatment are not met. The use of nonsense syllables as stimuli in investigations of that paradigm is, therefore, inadvisable.

Experiments I–IV also focus on the evaluation of two alternative experimental arrangements that have been used to assess the recovery of unlearned associations. In the conventional design, original learning (OL) and IL occur in immediate succession, and the length of the interval between the end of IL and the test of recall is varied. Absolute increases in recall as a function of the interval of delay provide direct evidence for recovery. The alternative procedure is to hold constant the length of the interval between the end of OL and the test of recall and to vary the point at which IL is introduced. The more closely IL follows OL, the greater is the proportion of the total interval during which recovery can occur. Which temporal point of interpolation (TPI) will produce the largest absolute rise in first-list recall will depend on the speed of recovery, on the one hand, and the rate of extraexperimental forgetting, on the other. In any event, if the interval is appropriate for the detection of recovery, recall should be lower when IL immediately precedes the test of retention than when it occurs directly after OL. For the A-C paradigm, however, no reliable differences in RI have been found as a function of the TPI, either when the test of first-list recall was paced (Archer & Underwood, 1951; Newton & Wickens, 1956) or when the MMFR procedure was used (Houston, 1967c). In all these cases the retention intervals were long—48 hr. in the first two studies cited and 1 wk. in the third. Thus, extraexperimental forgetting may have served to mask systematic effects of the TPI. For the C-D paradigm, Newton and Wickens did find progressive declines in first-list recall as the OL-IL interval was increased from 0 to 24 to 48 hr. As the investigators pointed out, the level of first-list recall was considerably higher in the C-D than in the A-C conditions, and the former may have been more sensitive than the latter to the influence of the temporal variable.

It was on the basis of these findings that Newton and Wickens (1965) introduced the hypothesis of generalized competition. The greater the separation between the OL and IL, the less recall will be influenced by the persistence of a set to give responses from the list learned last. It can be seen now that the concepts of recovery and of reversible response set interference generate the same predictions regarding the relation betwen RI and the TPI. According to the present analysis, spontaneous recovery, to the extent that it occurs, is reducible to the dissipation of response set interference. In light of these considerations, the influence of the TPI (immediately after OL vs. immediately before an MMFR test) was reexamined for the A-C and C-D paradigms, with a retention interval of the order of 20 min.[8] The studies were planned so as to permit a comparison of the effects observed by the two methods of assessing temporal changes in interference, i.e., by manipulation of the IL-test interval, on the one hand, and the TPI, on the other.

---

[8] Findings obtained by Keppel and Schwartz (1965) indicated that the TPI was likely to have an appreciable influence on RI for an interval of this length.

Experiment V extends the investigation of recovery to backward associations. It is known that the availability of backward associations is reduced when the paradigm of transfer entails an extinctive relation between such associations in the successive lists (e.g., Abra, 1967; Ellington & Kausler, 1965; Greenbloom & Kimble, 1965; Keppel & Underwood, 1962; Schwartz, 1968). The A-B, C-B paradigm provides a basic reference condition that defines the same relation for backward associations as does the A-B, A-C paradigm for forward ones. An attempt to apply the extinction analogy to the unlearning of backward associations at once raises difficult problems of interpretation. By definition, extinction requires the unreinforced elicitation of the critical response in the presence of the stimulus. There are apparent reasons for supposing that first-list forward associations are elicited during IL, but the same cannot be said about backward associations. It is by no means obvious what events during the acquisition of C-B might be responsible for the unlearning of B-A. Even a speculative answer to this question is difficult because the mechanism of associative interference under the C-B paradigm is not clearly understood. One possibility is that *S* uses backward associations to check correctness of responses before giving them (Twedt & Underwood, 1959). In the course of the checking process the intrusion of first-list stimuli may result in the false rejection of correct responses. Such sequences of events may lead in principle to the extinction or suppression of first-list stimuli. There is a troublesome discrepancy, however, between the weakness of the negative transfer effects and the massive amounts of RI in R-S recall. Associative interference is substantially less than for A-C, whereas the retroactive effects are, if anything, greater than those observed in forward recall in the latter condition (cf. Greenbloom & Kimble, 1965). It appears unlikely, therefore, that the mechanism responsible for the reduction of stimulus recall is closely linked to the occurrence of checking errors.

Another condition of R-S unlearning that may warrant consideration is bidirectional rehearsal of the paired associates during the acquisition of the successive lists. Wollen and Gallup (1965, 1968) have suggested that nominally backward associations function in the forward direction when *S* rehearses a pair two or more times in immediate succession, as he may well do in the course of a trial. Thus, rehearsal sequences like A-B-A-B and C-B-C-B may occur during first-list and second-list learning, respectively. Consequently, the conflicting associations B-A and B-C would be activated as *S* moves from the first to the second list. It is improbable, however, that such a state of affairs would be conducive to substantial amounts of associative interference and unlearning. In the transfer phase, *S* would be rehearsing a forward association to which he had been exposed just previously, with the R-S sequence occurring only as a by-product of repetition. Such intratrial rehearsal of currently correct associations should not be conducive to first-list stimulus intrusions.

There appear to be no compelling reasons for maintaining an a priori assumption that first-list backward associations are extinguished or suppressed in the transfer stage of the C-B paradigm. The possibility must be considered that the losses in R-S recall are entirely a test stage phenomenon. The key factor at the time of recall may be the change in function of the associated terms, with the B items becoming stimuli to which the A items must be given as responses. When this shift is first introduced, the stimuli from the list practiced last may be expected to be dominant and to inhibit the reproduction of those from the earlier list. On this assumption, response set interference which first develops at the time of the test would be responsible for the reduction in R-S recall. Thus, the classical conception of RI as entirely localized in the test stage may be valid for the case of R-S recall under the C-B paradigm.

It is not suggested that retroactive losses in the availability of first-list stimuli are always the result of test stage interference only. Different mechanisms may come into play depending on whether the conditions of IL entail the erroneous elicitation and consequent suppression of first-list stimuli. The latter conditions are probably met in the case of the A-B, B-C paradigm in which the first-list backward and the second-list forward associations are incompatible. The distinguishing characteristic of that paradigm is that the change in the function of the B items occurs during IL: the first-list responses are the stimuli in the second list. Once the B items become stimuli, they are likely to elicit the A terms associated with them; these interfering associations must be suppressed, and the newly prescribed C responses must be substituted for them. Now the amounts of loss in R-S recall under the C-B and the B-C paradigms are essentially the same (Cheung & Goulet, 1968), but according to the present analysis they may occur for different reasons. This interpretation would be supported if the temporal changes in interference were found to be different under the two paradigms. Since response suppression is assumed to be a reversible process, increases in R-S recall as a function of time are to be expected under the B-C paradigm; i.e., the tendency for B to elicit A, which was presumably activated during IL, should recover. Such recovery is not predicted for C-B since no comparable tendency was established during the acquisition of the transfer list. The conditions making for test stage interference should remain in force after an interval of time. That is, even when there is a delay between the end of IL and the test of R-S recall, the stimuli from the list learned last should retain the advantage of relative recency and interfere with the reproduction of stimuli from the earlier list. Put somewhat differently, recovery is to be expected only if the responses required on the test of recall were activated at some point during IL and then suppressed. This necessary condition of recovery is assumed to be met under the B-C but not the C-B paradigm. A failure to observe recovery under the latter condition was reported by Abra (1967) after a 24-hr. interval; there is no comparable information for the B-C treatment. In the present experiment, temporal changes in R-S recall under the C-B and B-C paradigms were determined over an interval of the order of 20 min.

Experiments I and II

Method

Design.—Experiments I and II, which were carried out concurrently, were designed to assess temporal changes in first-list recall under the A-B, A-C and the A-B, C-D paradigms. The interval between the end of IL and the test of recall was varied in Exp. I, and the TPI in Exp. II. With the two experiments performed simultaneously, there was some overlap of groups. For purposes of exposition the design of each experiment is first presented separately, and the overlapping groups are then identified.

The design of Exp. I comprised three conditions of interpolation (A-C, C-D, and rest) combined factorially with two intervals between the end of IL and the MMFR test of recall (2 and 18 min.). The end of OL and the beginning of IL were separated by 2 min. The two periods of delay between IL and test resulted in retention intervals for the first list of 14 and 30 min., respectively; these were, of course, the intervals after which the control groups were tested.

In Exp. II the interval between the end of first-list learning and the test of recall was kept constant at 30 min. Interpolated learning began either 2 or 18 min. after the last trial on the first list. The corresponding intervals between the end of IL and the MMFR test were 18 and 2 min., respectively. The three conditions of interpolation were again A-C, C-D, and rest. Since the retention interval for the first list was kept constant at 30 min., a rest group tested after that interval served as the common control condition for all experimental treatments. That control group and the two experimental groups for which the OL-IL interval was 2 min. were common to both experiments.

Since the assignment of *S*s to all conditions was concurrent, the use of overlapping groups does not influence the results of the separate experiments. There is, of course, a partial interdependence of the two sets of data which must be borne in mind when the findings are compared. A parallel evaluation of the relationship between the two types of experimental arrangements will also be made, however, in a further study.

Lists.—The experimental groups learned three successive lists conforming to either the A-B, A-C, A-D or the A-B, C-D, E-F paradigm. In the interest of brevity, these are referred to as the A-C and C-D paradigms. The control groups learned a single list. All lists consisted of eight paired associates, with single letters as stimuli and two-syllable adjectives as responses. There were three sets of eight stimulus letters and three sets of adjectives. The sets of letters were chosen so as to minimize acoustic similarity as far as possible both within and between lists; meaningful similarity of the responses was likewise kept as low as possible. Each of the three sets of stimulus letters was combined with each of the three sets of adjectives to yield nine different lists. The pairings were random, with the restriction that the stimulus letter and the first letter of the response word could not be the same. Successive responses to the same stimulus (A-B, A-C, A-D) always began with different letters. The nine lists were used as the first, second, and third task equally often. For purposes of assignment to the experimental treatments, the nine lists were grouped into three sets of three in accordance with the requirements of the transfer paradigm. Each of the six possible orders of the lists within a set was then assigned to 1 *S*. Thus the sequence of lists was different for each of the 18 *S*s within a group.

Learning procedure.—Learning was by the anticipation method at a 2:2-sec. rate, with a 4-sec. intertrial interval. There were four different random orders of presentation. The first list was learned to a criterion of 5/8. The two interpolated lists were presented for one study trial and six anticipation trials each. The time intervals between the end of OL and the beginning of IL in the two experiments have already been indicated. Practice on the third list always began 2 min. after the end of second-list learning. That part of the retention interval not occupied by IL under the experimental treatments, and the total interval under the control treatments, was filled with work on arithmetic problems.

Test of recall.—A variant of the MMFR procedure was used in the tests of recall. The *S* was handed an alphabetical list of all the stimuli used in OL and IL—8 letters in the A-B, A-C, A-D and the control conditions and 24 letters in the A-B, C-D, E-F conditions. The instructions to *S* were to recall orally all the pairs he had learned in any order he wished. The instructions emphasized that in giving a pair he was to call out both the stimulus and the response. In order to allow *S* to keep track of the frequency with which he had responded to each stimulus, he was requested to place a check mark next to the appropriate letter whenever he reproduced a pair. No feedback was given, except that *S* was informed when he repeated a pair. Such repetitions occurred rarely. The test was entirely *S*-paced. This procedure gave *S* maximal freedom to determine his order of output, but at the same time made it possible to obtain an exact record of the sequence in which pairs were recalled. After *S* had given his last oral response, he was handed his recall protocol and invited to make whatever changes or additions he wished (without obliterating the record of his oral responses). Space was provided for writing down responses which *S* was unable to pair with a particular stimulus. In the final phase of the test, experimental *S*s were asked to identify the list membership of all the responses they had recalled by placing the number 1, 2, or 3 next to each item on the protocol.

Subjects.—Taken together, Exp. I and II comprised eight (six experimental and two control) groups of 18 *S*s each. The *S*s were undergraduate students at the University of California who were not necessarily naive to verbal learning experiments. Assignment to conditions was in blocks of 8, with 1 *S* from each condition per block. The running order within blocks was determined by means of a table of random numbers.

Results and Discussion

First-list learning.—The mean number of trials to a criterion of 5/8 on List 1 was 5.11, with the means of individual groups ranging from 4.61 to 6.05. There were no reliable differences in speed of first-list learning, $F < 1$, for the comparison of all groups as well as for the separate comparisons of the groups included in Exp. I and II, respectively.

Interpolated learning.—In Exp. I the interval between the end of OL and the beginning of IL was kept constant at 2 min., whereas in Exp. II it was either 2 or 18 min. While there are some overlapping groups, it is useful to consider the transfer results separately.

In Exp. I the mean numbers of correct responses given in six trials of second-list learning were 23.86 and 28.86 for the A-C and C-D lists, respectively, $F\ (1, 70) = 7.52$, $p < .01$. Thus, there is a significant amount of negative transfer under the A-C relative to the C-D treatment. The corresponding values for third-list learning (Lists A-D and E-F) are 26.86 and 31.78, $F\ (1, 70) = 6.02$, $p < .02$. The difference between the two paradigms is the same as before. When Lists 2 and 3 are compared, the overall improvement in performance is found to be reliable, $F\ (1, 70) = 11.65$, $p < .01$. Since the amount of transfer remains constant, the Paradigm × Stage of Practice interaction is, of course, not significant, $F < 1$.

Table 1 contrasts second-list performance in immediate and in delayed IL (Exp. II). To permit an examination of the effects of the temporal variable on the course of ac-

TABLE 1

Mean Numbers of Correct Responses in Successive Two-Trial Blocks of Second-List Learning, Exp. II

TPI	Trials							
	A-B, A-C				A-B, C-D			
	1–2	3–4	5–6	Total	1–2	3–4	5–6	Total
Immediate	4.28	7.94	10.39	22.61	5.39	11.33	13.56	30.28
Delayed	5.39	10.00	11.56	26.95	6.44	9.50	12.33	28.27

quisition, the mean numbers of correct responses for successive blocks of two trials are presented. For the A-C paradigm, performance is consistently better in delayed than in immediate IL. The pattern for the C-D paradigm is more complex. Early in transfer learning the delayed group has an advantage, whereas on the later trials the difference is in the opposite direction. The net result is an increase in total scores as a function of the OL-IL interval for A-C and a smaller decrease for C-D. It is apparent that the influence of the TPI varies jointly with the paradigm and the stage of practice on the transfer task. The higher order TPI × Paradigm × Stage of Practice interaction is significant, $F\ (2, 136) = 21.54$, $p < .001$. For the total scores the TPI × Paradigm interaction falls short of significance, $F\ (1, 68 = 2.33$, $p > .05$. Since the changes for the two paradigms are in opposite directions, the overall effect of the TPI is not reliable, $F < 1$.

These findings suggest certain conclusions concerning temporal changes in the conditions of transfer under the two paradigms. It appears, first of all, that the effectiveness of A-B in interfering with the acquisition of A-C declines somewhat as a function of time, at least when the degree of OL is low. The *S*'s set to limit himself to the repertoire of B responses dissipates gradually; furthermore, the specific A-B associations may lose strength, even though the loss may not be great enough to be detected on an unpaced test of recall. The beneficial effect of delay in the A-C condition becomes understandable on these assumptions. The improvement for the first trial block in the C-D condition may likewise reflect the weakening of the response set established in OL. (The initial rises in performance also indicate that there was no substantial loss of warm-up during the interval.) But why should the delay of IL produce a decline in performance on the later trials of C-D learning? One plausible interpretation is that the discriminability between the stimulus terms in the successive lists is impaired as the OL-IL interval is lengthened, so that the functional paradigm becomes A′-D rather than C-D. The resulting interference becomes most pronounced as the first-list and second-list associations approach each other in strength. If this analysis is correct, the functional difference between the A-C and the C-D paradigm diminishes to some extent with the amount of delay between OL and IL.

There are no residual effects of the OL-IL interval on third-list learning, as would be expected if the last prior list is the major source of interference (cf. Postman, 1965; Underwood, 1945). In the C-D condition the mean total numbers of correct responses were 30.94 and 32.61 for immediate and delayed IL, respectively. The corresponding values for the A-C condition are 26.55 and 27.16. The TPI is not a significant source of variance nor does it interact with the paradigm of transfer, both $F\text{s} < 1$. When IL is delayed, the increase from List 2 to List 3 is greater for C-D than for A-C, whereas the amounts of improvement were closely similar in immediate IL. However, the higher order Stage of Practice × TPI × Paradigm interaction falls short of significance, $F\ (1, 68) = 2.87$, $p > .05$.

Recall: Effect of IL-test interval (Exp. I).—Figure 1 shows for all lists the mean numbers of responses recalled on the MMFR test. The stringent scores are shown, i.e., the numbers of correct responses given to the appropriate stimulus during the main phase of the test (prior to *S*'s written additions and corrections). On the immediate test the first-list scores reflect heavy RI for both paradigms. Contrary to what is usually observed after higher degrees of OL, the amounts of interference are approximately equal in the A-C and in the C-D condition; in fact, there is a slightly greater loss for the latter. The delayed test shows a rise in first-list recall for the C-D paradigm and a small decline for the A-C paradigm.

In order to adjust for variations in terminal OL performance, differences between the numbers correct on the last trial of OL and the recall scores were used in the statistical analysis of the retention data. The variation in terminal OL scores was relatively small and not significant ($F < 1$), with the means for the eight groups in Exp. I and II ranging from 5.39 to 5.61 and an overall mean of 5.46. Nevertheless, the use of difference scores adds some precision to the assessment of retention losses of individual *S*s. Thus, in the immediate control group the 10 *S*s who scored 5 on the criterial trial recalled 6.20 items, whereas the remaining 8 *S*s, who overshot the criterion, recalled 7.00 items. For the delayed control group the corresponding values are 6.00 ($n = 11$) and 7.00 ($n = 7$). In the analysis of variance of the first-list loss scores, the overall difference between the control condition and the combined experimental conditions is, of course, significant, F (1, 102) = 163.83, $p < .001$, i.e., there is a reliable amount of RI. The only other significant source of variance is the Paradigm (A-C vs. C-D) × Retention Interval interaction, F (1, 102) = 4.56, $p < .05$. This interaction reflects the fact that the temporal changes for the two paradigms are in opposite directions. In a further comparison of the C-D and control conditions the interaction with retention interval falls short of significance ($.10 > p > .05$).

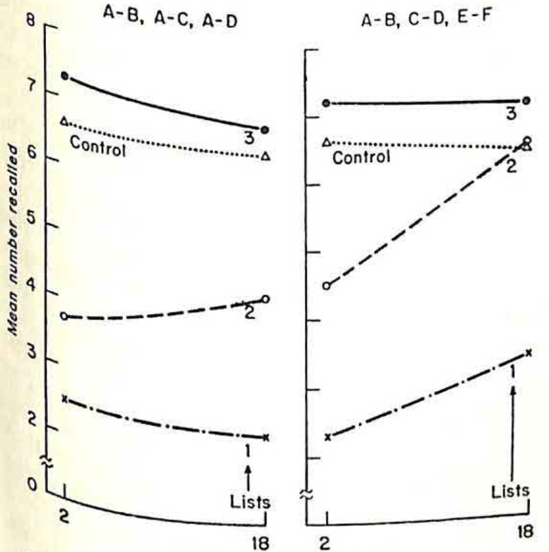


FIG. 1. Mean stringent recall scores on the immediate and on the delayed test, Exp. I. (To facilitate comparison with the experimental groups, the results of the common control groups are plotted in both panels.)

The recall scores fail to yield reliable evidence of recovery, although the C-D paradigm does show some gain on the delayed test. The possibility suggested itself that the present conditions were highly unfavorable to the observation of recovery since the first list was learned to a low criterion (5/8) and was followed by a substantial amount of IL on two different lists. A large proportion of the first-list associations might have been too weakly learned to regain availability even after some dissipation of response set interference. In fact, if such very weak items suffer some loss in strength during the retention interval, they might become more susceptible to the interference which remains effective after a period of delay. The divergence in the temporal trends for the two paradigms becomes understandable on the assumption that the initial degree of response suppression, and hence the residual interference on the delayed test, is less in the C-D than in the A-C condition.

To test the implications of this hypothesis, the recall scores were examined for items varying in speed of acquisition during first-list learning. For this purpose the pairs in each *S*'s list were ranked from lowest (1) to highest (8) according to the number of correct anticipations given during the acquisition trials. Figure 2 shows the percentages of correct first-list responses (stringent scoring) for successive blocks of two ranks each.

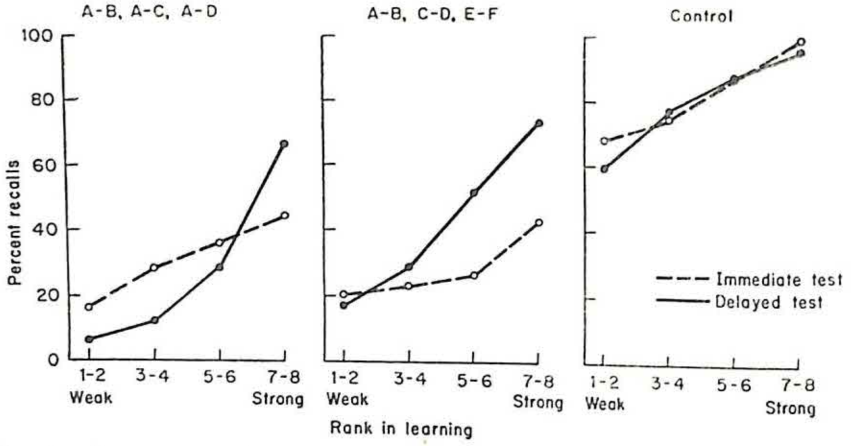


FIG. 2. Percentages of recalls on the immediate and on the delayed test as a function of the strength of the items in acquisition, Exp. I.

It is apparent that in the A-C condition the temporal changes in recall depend on item strength: for the relatively weak items, represented by Ranks 1–6, there is a decline between the immediate and the delayed test, whereas for the two strongest items (Ranks 7–8) there is a substantial rise. The amount of recovery also increases with item strength in the C-D condition, but the minimal degree of OL sufficient for recovery is lower than for A-C. The difference between the two paradigms in the relation between item strength and recovery is reflected in the net increase for C-D and the net decrease for A-C.

An analysis of variance was carried out on the weighted loss scores for Ranks 1–6 and 7–8. The control groups were not included in this analysis since their loss scores for strong items had essentially zero variance. The Item Strength × Retention Interval interaction is significant, F (1, 68) = 9.79, $p < .01$; the higher order interaction of the latter two variables with paradigm of transfer is not significant, $F = 1.78$. For Ranks 7–8 the increase between the immediate and the delayed test is reliable ($p < .01$). Thus, there is recovery for the strongest items under both paradigms. It is recognized, of course, that this analysis is post hoc since the division between strong and weak items was made after inspection of the data. The test is useful, however, in confirming the visual impression. More important, the implications of the analysis were subjected to an independent test in a subsequent experiment. In the meantime, two conclusions suggested by these results deserve emphasis: (*a*) Under given conditions of IL, only those items whose strength at the end of OL exceeds a critical value are likely to show recovery; and (*b*) when retention losses are heavy, equal amounts of immediate RI do not necessarily reflect equal amounts of unlearning. A subsequent test of recovery may be sensitive to differences in interference that are not detectable on an immediate test when recall scores are near the floor of performance. In the present case the divergence of the scores on the delayed test suggests that the interference was actually heavier in the A-C than the C-D condition.

Second-list recall is considered next. On the immediate test the number of correct responses is higher for C-D than for A-C, as would be expected on the basis of the difference in speed of transfer learning. There are rises in second-list recall on the delayed tests under both paradigms. For C-D the mean loss score declines from 1.00 to .22, and for A-C it declines from 2.39 to 1.33. (The loss scores again correct for sampling fluctuations in the terminal level of List 2 learning. With this correction, the greater recovery for C-D suggested in Fig. 1 is eliminated. It should be noted that retention scores are highly sensitive to relatively minor differences in degree of learning when a fixed number of trials is given in acquisition.) Analysis of variance of the loss scores shows the main effects of paradigm and of retention interval to be significant, $F\ (1,\ 68) = 5.54,\ p < .05$; and $F\ (1,\ 68) = 10.30,\ p < .01$, respectively. There is no reliable Paradigm × Retention Interval interaction, $F < 1$. Thus, the second list appears to show a significant amount of recovery under both paradigms. In the absence of appropriate control groups, this conclusion must remain tentative, although it is very unlikely that comparable rises in recall would occur without interpolation of the third list. There are only slight changes in third-list recall, and these are not reliable. It is interesting to note, however, that the loss between the immediate and the delayed test is somewhat greater for the A-C than the C-D condition. The present conditions are clearly not favorable to the observation of proactive effects since the third list was learned to a higher degree than the preceding ones.

The lenient scores, based on the number of responses recalled regardless of correctness of pairing, are not reported since they provide little new information. The differences between lenient and stringent scores were comparable for the various experimental treatments. Thus, the pattern of results does not change appreciably when lenient rather than stringent scores are considered.

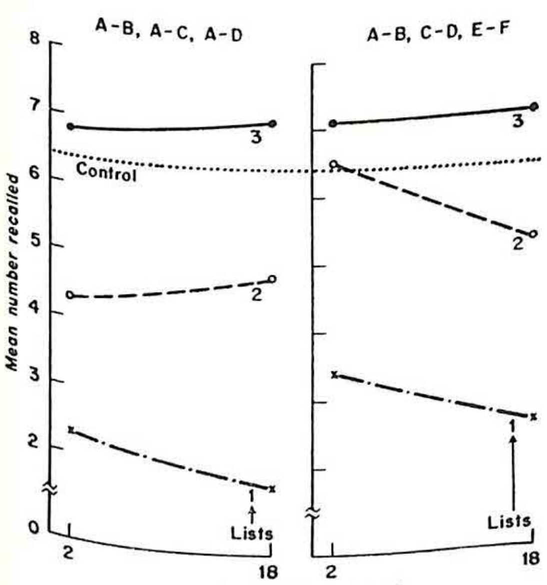


FIG. 3. Mean stringent recall scores after immediate and after delayed IL, Exp. II.

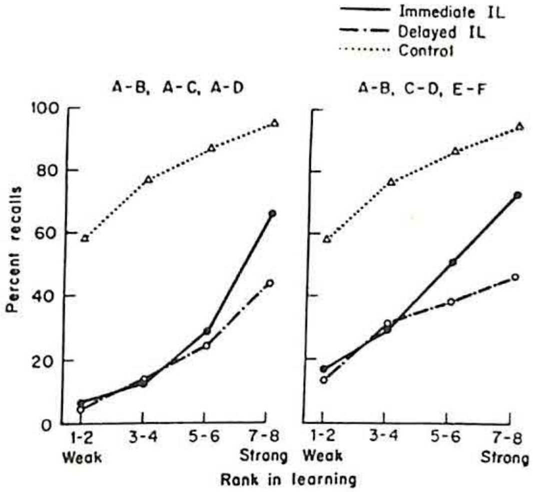


FIG. 4. Percentages of recalls after immediate and after delayed IL as a function of the strength of the items in acquisition, Exp. II.

Recall: Effect of TPI.—Figure 3 shows the influence of the TPI on the stringent recall scores. Under both paradigms, first-list recall is lower when IL occurs at the end rather than at the beginning of the retention interval. While the decreases are relatively small, analysis of variance of the loss scores shows the main effect of the TPI within the experimental treatments to be significant, $F\ (1,\ 85) = 5.73,\ p < .02$. Since recall is higher in the C-D than in the A-C condition after a period of delay, there is also a significant overall difference between the two paradigms, $F\ (1,\ 85) = 8.96,\ p < .01$. Paradigm does not interact, however, with TPI, $F < 1$. Recall as a function of rank in acquisition is plotted for the relevant groups in Fig. 4. Comparison with the corresponding functions in Fig. 2 makes it clear why the total recall scores change significantly with the TPI but not with the IL-test interval. For the A-C paradigm, performance on the test immediately following IL is lower in Exp. II than in Exp. I. Whatever loss the weakest items suffer during the interval occurs for both groups tested after a delay; moreover, there is better acquisition of the interfering second list in delayed than in immediate IL. Consequently, in the evaluation of the difference attributable to the TPI, the increases for strong items are not offset by losses for weak items. For the C-D paradigm, recall on the test immediately after IL tends to be higher at the end than at the beginning of the retention interval, probably because the interfering second list is learned less well. This difference is present for all but the strongest items. However, the level of recovery in this condition is sufficiently high for an overall effect of the TPI to remain in evidence.

There is little change in second-list recall as a function of the TPI in the A-C condition, while there is a difference in the expected direction in the C-D condition. Analysis of variance of the loss scores does not yield any significant effects. Since the OL-IL interval had at least some effect on speed of second-list learning, and in opposite directions for the two paradigms, a precise evaluation of the relative amounts of second-list recall is difficult in any case. Variations in third-list recall are minor and not reliable. The slight rises that are in evidence probably reflect the reduction in the length of the retention interval for the interpolated lists when TPI is delayed. As before, lenient scores are not considered and for the same reasons.

Order of output in recall.—One of the characteristic temporal changes in performance observed in earlier studies (Postman et al., 1968) is a systematic shift in the order of output of the available responses. The first list was found to move forward in the recall sequence as the interval between the end of IL and the test was lengthened. For purposes of examining the relative priority of the three lists in recall, the responses reproduced by each *S* were ranked according to the order in which they were given by him during the main phase of the test. Incorrectly as well as correctly paired responses were included since the question of interest concerns the list membership of the responses rather than the integrity of specific associations. The ranks were then converted into *C* scores (Guilford, 1950, pp. 302 ff.). The *C* scale, which is based on centile ranks, has a mean of 5 and a range from 0 (lowest) to 10 (highest). Thus comparable measures were obtained for *S*s recalling different numbers of items. The *S*'s mean *C* score was then determined for each of the three lists. The *S*s who failed to recall a single response from one or more of the lists were omitted from the analysis. In order to minimize differential *S* selection, for each *S* eliminated on this basis, one *S* matched as closely as possible in recall performance to the one rejected was removed from each of the remaining groups as well. This procedure resulted in the elimination of three *S*s from each of the groups, leaving $n = 15$ per group.

The means of the mean *C* scores are summarized in Table 2. The effect of the IL-test interval (Exp. I) is considered first. On the immediate test the priority of the list learned last is clearly apparent. In the C-D condition, order of output is directly related to relative recency. In the A-C condition, there is a suggestion of a V-shaped "serial position effect," although the third list has a clear advantage. A shift in the expected direction is present on the delayed test for the C-D paradigm, i.e., a reduction of the bias in favor of recently learned responses. On the other hand, the picture does not change appreciably on the delayed test for the A-C condition. This difference between the two paradigms is consistent with the fact that there was an absolute rise in total first-list recall only for C-D. Since the scores for the three lists are not independent, statistical analysis is limited to the first-list measures. These scores were chosen because interest centers on first-list recovery and all retention intervals were defined with reference to the first list. There are no significant changes in these values. This outcome is not surprising in view of the absence of recovery for A-C and the marginal amount of improvement in recall for C-D.

The effect of the TPI on the order of output (Exp. II) is now considered. The relevant measures are given in Table 2. There is a shift in output order corresponding to the reliable decline in first-list recall with a delay in the TPI. This shift encompasses the scores for all lists in the C-D condition, but is limited to the first two in the case of A-C. In the analysis of the first-list scores, TPI is a significant source of variance, *F* (1,

TABLE 2

Relative Priority of Responses from the Three Lists in Order of Output, Exp. I and II

TPI	Retention interval	A-B, A-C			A-B, C-D		
		List 1	List 2	List 3	List 1	List 2	List 3
Immediate	Immediate	4.69	4.46	5.45	4.08	4.80	5.74
Immediate	Delayed	4.65	4.34	5.61	4.56	5.12	5.09
Delayed	Delayed	4.30	4.61	5.55	3.65	4.70	5.71

Note.—Entries are means of mean *C* scores.

56) $= 4.82$, $p < .05$, but does not interact reliably with paradigm, $F < 1$.

It appears that variations in output order as a function of relative recency are less pronounced in the A-C than the C-D condition. One possible reason is that in the former but not in the latter case the recall sequence is constrained by *S*s' tendency to give in succession all the available responses to a given stimulus. To assess the strength of this tendency for the A-C groups, the degree to which responses to the same stimulus clustered in recall was determined. A clustering index recommended by Dalrymple-Alford[4] was used:

$$C = \frac{R - \min R}{\max R - \min R},$$

where R refers to the number of repetitions of responses from the same class (in the present case defined by a common stimulus), with max R and min R designating the maximal and minimal possible number of repetitions, respectively. This index is computed separately for expected and for observed values. The difference between the two ratios (D) provides a measure of the amount of clustering in excess of chance. When IL is immediate, the value of D for the immediate test is .35 and for the delayed test is .14. On the test after delayed IL the index rises again to .36. All three values differ significantly from zero, the lowest one at the .05 level and the two higher ones at the .01 level. Thus, a significant tendency to group the responses to the same stimulus is present on all tests. The critical factor determining the amount of clustering seems to be the IL-test interval. Orthogonal comparisons show that the index is significantly higher ($p < .05$) for the two tests following IL immediately than for the test separated from IL by a period of delay. The results for the former two tests are virtually identical, i.e., the first-list retention interval has no effect. The decline in the amount of clustering suggests that immediately after the end of IL, *S* is disposed to respond to each stimulus in turn, just as he did during the preceding acquisition trials, but that this disposition is weakened as a function of time. After a delay, *S* is more likely to scan available responses and to match them with the stimuli before him. Consequently there would be a reduction in the tendency to give all the responses to a given stimulus before proceeding to the next one, and the probability of recurrent attempts at recall to the same stimulus would increase.

List identification.—Table 3 presents the percentages of correct identifications of the list membership of the responses. Only appropriately paired responses given during the main phase of the test were used in this tabulation. In the interest of obtaining stable values, the percentages are based on the pooled responses of each group. For each list the percentage of false alarms is also shown, e.g., the percentage of items from Lists 2 and 3 assigned to List 1, etc. The differences between the percentages of correct identifications and of false alarms may be considered as measures of accuracy corrected for guessing. These measures uniformly reflect the positive effects of primacy and recency, i.e., the first and the last list have an advantage over the middle one. This pattern agrees with that reported by Keppel

---

[4] E. C. Dalrymple-Alford, personal communication, 1969.

TABLE 3

MEASURES OF ACCURACY OF LIST IDENTIFICATION, EXP. I AND II

TPI	Retention interval	List 1			List 2			List 3		
		% CI	% FA	Difference	% CI	% FA	Difference	% CI	% FA	Difference
A-B, A-C										
Immediate	Immediate	84.4	6.6	77.8	77.6	8.0	69.6	91.6	5.4	86.2
Immediate	Delayed	78.0	9.5	68.5	69.2	12.3	56.9	87.7	7.6	80.1
Delayed	Delayed	90.3	5.1	85.2	75.9	11.3	64.6	88.3	8.5	79.8
A-B, C-D										
Immediate	Immediate	63.4	7.6	55.8	65.4	17.0	48.4	86.0	12.3	73.7
Immediate	Delayed	59.7	9.0	50.7	54.7	30.5	24.2	68.0	21.8	46.2
Delayed	Delayed	74.4	9.7	64.7	58.3	23.2	35.1	73.8	17.5	56.3

Note.—CI = Correct identifications; FA = false alarms.

(1967) and points to the effectiveness of distinctive temporal "tags" as cues to list differentiation (cf. Underwood & Ekstrand, 1967). The highest percentages of incorrect assignments tend to be to List 2, in accord with the usual finding that failures of differentiation are most common between adjacent lists. The introduction of an IL-test interval (Exp. I) leads to a decline in the accuracy of the identifications. In the A-C condition the losses do not vary appreciably for the three lists; in the C-D condition the degree of impairment is clearly less for List 1 than for Lists 2 and 3. A limited amount of recovery and corresponding shifts in the order of output were observed for the same condition. With the first-list retention interval held constant, delay of IL (Exp. II), in general, favors list differentiation. Again, temporal cues, in this case related to the separation between List 1 and Lists 2 and 3, appear to be a source of facilitation. Finally, it should be noted that the level of accuracy is markedly lower for the C-D than the A-C groups. At a low level of learning, the C-D paradigm is conducive to errors based on confusion of stimulus terms as well as of response terms. Only the latter source of error is present under the A-C paradigm.

EXPERIMENTS III AND IV

The results of Exp. I and II yielded only suggestive information regarding the relation between first-list recovery and the conditions of transfer. Whatever trends were present indicated greater recovery for C-D than for A-C, contrary to the implications of the extinction model. However, in the absence of significant increases in total recall scores, no firm conclusions can be drawn from these findings. The next step was suggested by the results of the item analyses, which indicated that a higher degree of OL than that used in Exp. I and II is required for the detection of substantial amounts of recovery. The purpose of Exp. III and IV was to verify this implication and to permit a further examination of the variables to which the previous study was addressed.

Method

Design and procedure.—The design, materials, and procedure were exactly the same as in the earlier study, with Exp. III and IV corresponding to Exp. I and II, respectively. The only change was that first-list learning was carried to a criterion of $5/8 + 4$ additional trials. The additional practice trials beyond the previous criterion were expected to yield distributions of item strengths that would be favorable to the occurrence of significant amounts of recovery.

Subjects.—There were again 18 *S*s in each of the eight groups, drawn from the same population and assigned to conditions in the same manner as before.

Results and Discussion

First-list learning.—The grand mean of the number of trials to a criterion of 5/8 was 4.70 (as compared to 5.11 in Exp. I and

TABLE 4

MEAN NUMBERS OF CORRECT RESPONSES IN SUCCESSIVE TWO-TRIAL BLOCKS OF SECOND-LIST LEARNING, EXP. IV

TPI	Trials							
	A-B, A-C				A-B, C-D			
	1–2	3–4	5–6	Total	1–2	3–4	5–6	Total
Immediate	5.39	9.78	12.83	28.00	7.50	11.82	14.22	33.54
Delayed	5.61	9.83	11.78	27.22	5.50	10.44	12.50	28.44

II). The means of the individual groups ranged from 4.22 to 5.28, $F < 1$. The mean number correct on the terminal trial of first-list learning was 6.45, with a range of group means from 6.22 to 6.67, $F < 1$. The corresponding mean obtained previously was 5.46. While the additional practice produced only a moderate change in the level of terminal performance in OL, the average number of correct anticipations per item was increased substantially.

Interpolated learning.—When IL was immediate (Exp. III), the mean numbers of correct responses in six trials of second-list learning were 27.89 for A-C and 32.78 for C-D, $F\ (1, 68) = 7.36$, $p < .01$. The level of second-list learning is appreciably higher than in Exp. I. This rise in performance may be attributed to an increase in nonspecific transfer effects at the higher level of first-list practice. The absolute amount of negative transfer is identical with that in Exp. I. The mean numbers correct in six trials of third-list learning were 28.94 and 32.92 for the A-C and C-D paradigms, respectively, $F\ (1, 68) = 4.60$, $p < .05$. With nonspecific transfer effects leveling off rapidly, there is no longer a significant improvement between the second and the third list.

The measures of second-list performance in immediate and in delayed IL (Exp. IV) are presented in Table 4. In order to permit a comparison with the corresponding results in Exp. II (see Table 1), the mean numbers of correct responses are shown for successive blocks of two trials each. There is little difference between the two TPIs for the A-C paradigm, whereas performance in the C-D condition is consistently lower in delayed than in immediate IL. While the TPI × Paradigm interaction is not significant ($F = 1.21$), it is important to note that as a result of the differential temporal changes, the absolute amount of negative transfer in delayed IL is minimal. In the earlier study, A-C performance was found to increase and C-D learning to decrease late in learning as the OL-IL interval was lengthened. It was suggested that the effectiveness of A-B in interfering with A-C declined as a function of time, whereas in the case of C-D an impairment of stimulus discriminability resulted in associative interference. The latter source of negative transfer was assumed to become increasingly effective when first-list and second-list learning approach each other in strength. The same considerations apply to the present results when account is taken of the fact that first-list learning was carried to a higher level and second-list learning progressed more rapidly. As a consequence, the decline in the interfering potential of A-B was not sufficiently great to result in improved A-C performance, and associative interference came into play early during the acquisition of C-D.

There are no residual effects of the TPI on third-list learning. For the A-C paradigm the mean total numbers of correct responses were 28.44 and 29.72 in immediate and in delayed IL, respectively. The corresponding values for the C-D paradigm are 32.44 and 31.38. There is no significant improvement between the second and the third list.

Recall: Effect of IL-test interval (Exp. III).—Figure 5 shows the mean stringent scores for the two retention intervals. The

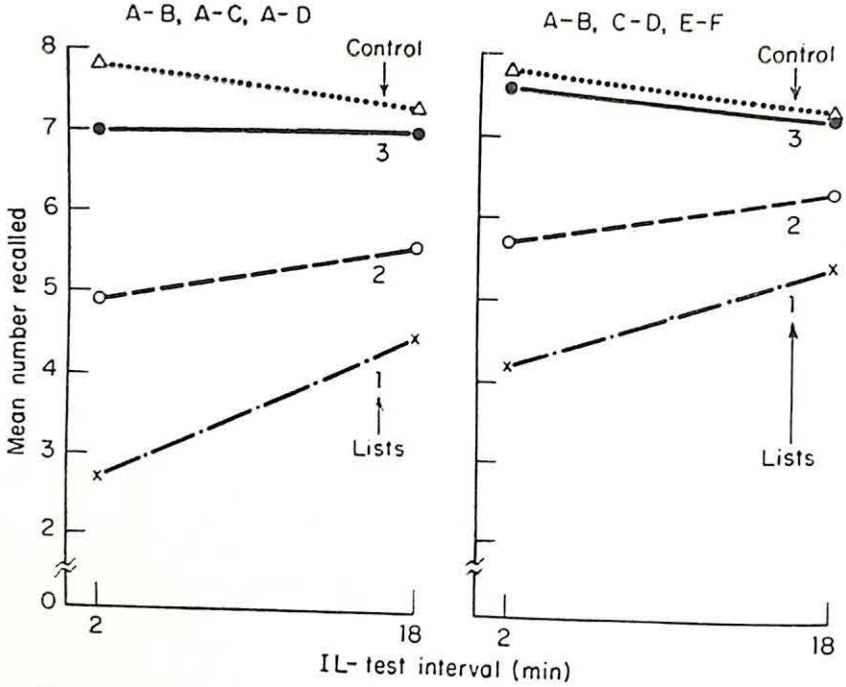


FIG. 5. Mean stringent recall scores on the immediate and on the delayed test, Exp. III. (To facilitate comparison with the experimental groups, the results of the common control groups are plotted in both panels.)

first-list scores again yield substantial RI for both paradigms, although recall is somewhat higher for C-D than for A-C. As expected, there are clear rises in first-list recall on the delayed test. The amounts of recovery are comparable under the two conditions of transfer. The decrease in loss scores is 1.4 for A-C and 1.6 for C-D. In the control condition the amount forgotten increases by .5 items. In the analysis of variance of the loss scores, F (1, 102) = 79.66 for the difference between the control condition and the combined experimental conditions. There is also a significant overall difference between A-C and C-D in favor of the latter, F (1, 102) = 4.73, $p < .05$. The critical evidence for recovery is provided by the presence of a significant Control–Experimental Difference × Retention Interval interaction, F (1, 102) = 5.80, $p < .02$. Within the experimental treatments, paradigm does not interact with retention interval, $F < 1$.

It was found in the earlier study that after a low degree of OL, only the relatively strong first-list items showed recovery. As Fig. 6 shows, recovery occurs over a wider range of item strengths when the degree of OL is increased. In the A-C condition the function for the delayed test lies above that for the immediate test at all ranks. In the case of C-D the functions diverge only for the stronger half of the items. However, as the figure indicates, recall of the weaker items on the immediate test was relatively high under this condition of transfer, perhaps because of the use of a fixed number of practice trials beyond criterion. In any event, the opportunities for recovery of the weaker items were more limited than under the A-C treatment.

Second-list recall increases between the immediate and the delayed test for both paradigms. Analysis of variance of the loss scores shows the main effect of retention interval to be significant, F (1, 68) = 4.87, $p < .05$. As would be expected on the basis of the trends in acquisition, the loss scores are greater for A-C than for C-D on the immediate test, but this difference is largely eliminated after a delay. However, the Paradigm × Retention Interval interaction is not

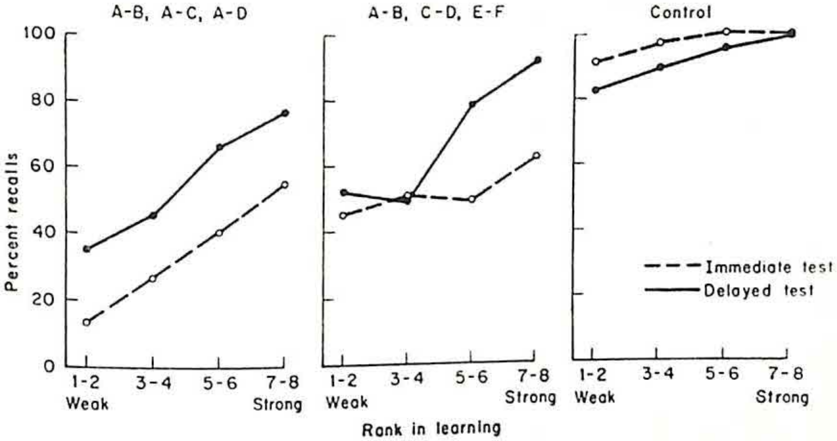


FIG. 6. Percentages of recalls on the immediate and on the delayed test as a function of the strength of the items in acquisition, Exp. III.

reliable, $F = 2.00$. There are no significant changes in third-list recall. As in the previous study, lenient scores are not reported since they add no essential new information.

Recall: Effect of TPI (Exp. IV).—As Fig. 7 shows, first-list recall drops substantially under both conditions of transfer when the TPI is moved from the beginning to the end of the retention interval. The difference in the amount of loss is significant, $F\ (1, 85) = 10.28$, $p < .01$. The drops in performance for A-C and for C-D are essentially equal, with $F < 1$ for the Paradigm × TPI interaction. Immediate IL produces higher recall than delayed IL over the entire range of item strengths under both paradigms (see Fig. 8). It is useful to note that manipulation of the TPI produces larger changes in recall than the introduction of an IL-test interval. Parallel comparisons

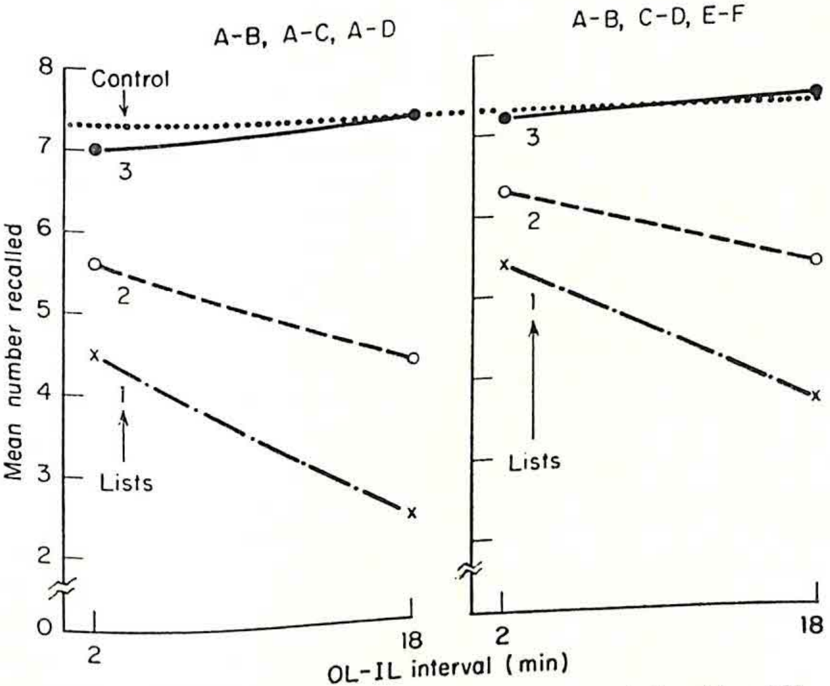


FIG. 7. Mean stringent recall scores after immediate and after delayed IL, Exp. IV.

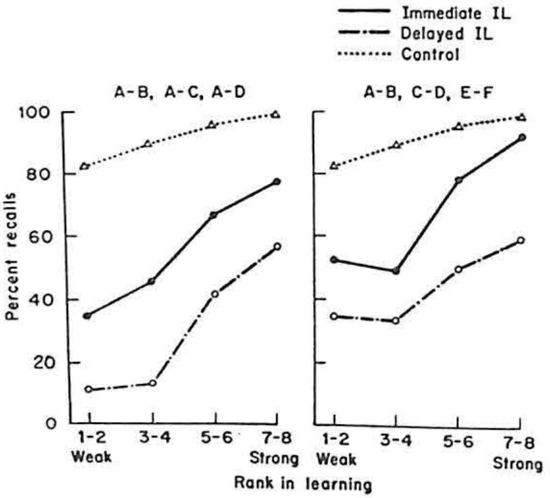


FIG. 8. Percentages of recalls after immediate and after delayed IL as a function of the strength of the items in acquisition, Exp. IV.

were difficult to make in the first study because of the generally low level of recall, but the same seemed to be true for the A-C though not for the C-D paradigm.

The second-list scores also show a decline when IL is delayed rather than immediate. The effect of the TPI fails, however, to reach significance, $F = 2.84$, nor does it interact with paradigm, $F < 1$. Delay of IL again produces small and nonsignificant increases in third-list recall attributable to the reduction in the retention interval for that list.

Order of output in recall.—Table 5 presents the C scores indexing the relative priority of the three lists in the order of output. The scores were calculated in the manner described previously. Two Ss were eliminated from each group on the same basis as before, so that $n = 16$ per group. In general, the degree of priority in recall is directly related to the recency of acquisition. The systematic effects associated with the temporal changes in interference are apparent. Recovery (Exp. III) is accompanied by a shift forward in the recall sequence of the first list and, to a lesser extent, of the second list, with a correlated reduction of the bias in favor of the most recent list. For the first-list scores the main effect of the retention interval is significant, $F\ (1, 60) = 6.00$, $p < .02$. There is no Paradigm × Retention Interval interaction, $F < 1$. It should be noted that on both tests the first-list scores are higher for C-D than for A-C, suggesting more effective suppression of the original list under the latter than under the former condition. The overall difference between the paradigms is significant, $F\ (1, 60) = 5.93$, $p < .02$. As expected, the direct relation of priority to recency of acquisition becomes more pronounced when IL is delayed rather than immediate (Exp. IV). The effect of the TPI is significant, $F\ (1, 60) = 4.82$, $p < .05$. Again there is no reliable interaction of the temporal variable with paradigm, $F < 1$.

In spite of the absence of significant interactions, there is some indication, as there was in the earlier study, that the shifts in the order of output are less sharp in the A-C than in the C-D condition. It was suggested that Ss' tendency to cluster the responses to a common stimulus imposes a certain amount of constraint on variations in the order of output. In the present results there is again substantial evidence of such clustering. When IL is immediate, the index is .36 on the immediate test and .27 on the delayed test. With delayed IL the value is .38. All three indexes are significantly greater than zero ($p < .01$); while they do not differ

TABLE 5

RELATIVE PRIORITY OF RESPONSES FROM THE THREE LISTS IN ORDER OF OUTPUT, EXP. III AND IV

TPI	Retention interval	A-B, A-C			A-B, C-D		
		List 1	List 2	List 3	List 1	List 2	List 3
Immediate	Immediate	3.98	4.90	5.54	4.37	4.92	5.53
Immediate	Delayed	4.37	5.06	5.26	4.94	4.98	5.14
Delayed	Delayed	4.02	4.85	5.55	4.04	4.61	5.84

Note.—Entries are means of mean C scores.

TABLE 6

Measures of Accuracy of List Identification, Exp. III and IV

TPI	Retention interval	List 1			List 2			List 3		
		% CI	% FA	Difference	% CI	% FA	Difference	% CI	% FA	Difference
					A-B, A-C					
Immediate	Immediate	87.8	5.5	82.3	75.3	8.5	66.8	91.4	10.1	81.3
Immediate	Delayed	80.2	6.1	74.1	72.3	15.4	56.9	84.2	9.3	74.9
Delayed	Delayed	86.4	2.9	83.5	84.6	8.0	76.6	93.1	6.5	86.6
					A-B, C-D					
Immediate	Immediate	81.3	6.3	75.0	60.2	17.5	42.7	81.6	15.7	65.9
Immediate	Delayed	82.6	7.4	75.2	62.3	21.9	40.4	67.7	15.1	52.6
Delayed	Delayed	73.4	5.2	68.2	66.7	22.2	44.5	79.1	13.1	66.0

Note.—CI = Correct identifications; FA = False alarms.

reliably from each other, the introduction of an IL-test interval again leads to some reduction in clustering. It would appear that the disposition to respond to each stimulus in turn in accordance with the requirements of the acquisition procedure persists longer as the degree of OL increases.

List identification.—The measures of list identification, which are presented in Table 6, are considered next. The primacy and recency effects favoring accurate identification of the first and third lists are in evidence throughout. As before, the scores are higher for A-C than for C-D. With only one exception, incorrect assignments are made most frequently to List 2. The IL-test and the OL-IL intervals have little effect on the accuracy of list identification, i.e., the changes observed at a lower level of learning are no longer present. In a previous investigation (Postman et al., 1968) it was found that rises in first-list recall were accompanied by increased accuracy of list identification. In that study the lists were longer and the amount of RI greater than in the present one; there was evidence of proactive interference in the recall of the third list as well. These facts, in conjunction with the trend observed in Exp. I, suggest that there may be a direct relation between recovery and accuracy of list identification only under conditions of heavy interference in which both the original and the interpolated lists are at relatively low levels of strength. In any event, changes in the measures of list identification appear to be less closely and less reliably associated with recovery from interference than are systematic shifts in the order of output of the responses. This fact may be viewed as consistent with the assumption that recovery reflects primarily the dissipation of response set interference.

Experiment V[5]

Method

Design.—In Exp. V, interest centered on the recovery of backward associations as a function of the conditions of transfer. Two paradigms of transfer were compared, viz., A-B, C-B and A-B, B-C. Recovery was expected only under the latter treatment for the reasons discussed in the introduction. The design comprised the factorial combinations of three treatments and two retention intervals. The treatments included the two transfer paradigms and a control condition in which *S*s learned and recalled a single list. The first-list retention interval was either 16 or 36 min. In the experimental conditions these intervals represented delays between IL and test of 2 and 22 min., respectively.

Lists.—The learning materials were lists of eight paired associates, with two-syllable adjectives as both stimuli and responses. The inclusion of the A-B, B-C paradigm made it necessary to draw stimuli and responses from a common pool of items. Three sets of eight adjectives each were used in the construction of the lists, with each set occupying the stimulus and the response positions equally often. There were six different lists of paired associates representing the six possible permutations of the three sets of words. These six lists were used equally often as the first and as the second task. The permissible combinations of first and second lists were determined by the requirements of the transfer paradigm. Similarities of meaning and pronunciation were minimized both within pairs and between lists.

---

[5] The authors are grateful to Ellen Rossman for her assistance.

Learning procedure.—Learning was by the anticipation method at a 2:2-sec. rate, with a 4-sec. intertrial interval. There were four different orders of presentation. The first list was learned to a criterion of 7/8, and there were 15 anticipation trials on the second list. Practice on the second list began 2 min. after the end of first-list learning. The instructions for second-list learning included an explanation of the relation between the two lists. A visual display, in which simple geometric forms were substituted for the conventional letter designations of stimulus and response terms, was used to illustrate the nature of the paradigm. These additional instructions were introduced primarily for the purpose of minimizing the confusion likely to follow the change in the function of the B terms in the A-B, B-C condition; the appropriate information was, of course, given under both treatments. The part of the retention interval not occupied by IL in the experimental conditions, and the entire interval in the control condition, was filled with work on arithmetic problems.

Test of recall.—In view of the dual function of the B term in A-B, B-C, it was considered best to test R-S recall for the first list only. The first-list responses were presented in the window of the memory drum one at a time. The items were presented in the order that would have occurred if there had been an additional learning trial. The *S*s were instructed to call out the stimulus that had been associated with a given response during first-list learning. The test was entirely *S*-paced. At the end of the test, *S* was permitted to make written changes and additions on his protocol and was encouraged to include stimulus terms that he was unable to pair with the appropriate response.

Subjects.—The *S*s were undergraduate students at the University of California. There were 18 *S*s in each of the six groups. Assignment to conditions was in blocks of 6, with 1 *S* from each condition per block. The running order within blocks was determined by means of a table of random numbers. In order to restrict the range of variation in ability, *S*s who reached criterion on the first list in 1 trial or who required more than 15 trials were discarded and replaced by the next *S* appearing in the laboratory.

Results and Discussion

First-list learning.—The mean number of trials to a criterion of 7/8 was 7.32, with the means of individual groups ranging from 5.72 to 7.89, $F = 1.25$. On the criterial trial the mean number correct was 7.14, with a range of group means from 7.06 to 7.22, $F < 1$.

Second-list learning.—The mean numbers of correct responses in 15 trials of anticipation learning were 99.9 for the C-B paradigm and 90.6 for the B-C paradigm. This difference is significant, $F\ (1, 70) = 6.32$, $p < .01$. Inspection of the acquisition curves shows a substantial separation between the two conditions of transfer on the first 6 trials, with a gradual convergence thereafter. While the absolute amounts of paradigm-specific transfer cannot be determined in the absence of a C-D control base line, it can be concluded that the relative degree of associative interference is greater in the B-C than in the C-B paradigm. The apparent source of interference in the B-C condition is the change in the function of the B terms. The frequency of stimulus errors, i.e., of stimulus terms given as responses, is higher for the B-C than for the C-B condition, although the rates are low in both cases. The mean percentage (to the base of opportunities) of stimulus errors for B-C is 4.8 and for C-B is 1.7. The difference approaches significance by the Mann-Whitney test, $z = 1.80$, $.10 < p < .05$. A complementary finding is that the mean percentage of misplaced responses is somewhat higher (22.0) for C-B than for B-C (18.4), but this difference is not reliable. While the trends in error rates are not conclusive, they are consistent with the assumption that the elicitation of stimulus terms as responses occurs with differential frequency under the two paradigms. Consequently this error tendency may be more thoroughly suppressed under the B-C than the C-B treatment, and this inhibitory set may generalize to the stimulus terms of the first list at the time of R-S recall. In addition, as has already been suggested, the first-list stimulus terms themselves are subject to suppression to the extent that they occur as intrusions during the acquisition of the transfer list. Such overt intrusions were, as usual, infrequent, but the total number of occurrences was greater for B-C (10 cases) than for C-B (3 cases).

TABLE 7

MEAN NUMBERS OF STIMULUS TERMS RECALLED IN MFR, EXP. V

Retention interval	A-B, C-B		A-B, B-C		Control	
	Stringent	Lenient	Stringent	Lenient	Stringent	Lenient
Immediate	2.72	3.50	2.72	3.50	5.72	6.11
Delayed	2.17	2.89	3.61	4.33	5.44	5.67

Recall.—Performance on the test of recall is summarized in Table 7. The stringent scores represent correctly paired stimulus terms recalled during the main phase of the test. In lenient scoring, credit was given for all stimulus terms recalled at any time during the testing procedure regardless of whether they were paired with the appropriate response. The lenient as well as the stringent scores are presented because the differences between them are appreciable and serve to call attention to the relative weakness of specific R-S associations at the end of forward learning.

The stringent scores yield a substantial amount of RI on the immediate test. The retention losses are identical for the two paradigms. On the delayed test, however, there is a decline for C-B and a rise for B-C. The control condition shows a moderate loss between the immediate and the delayed test. Figure 9 shows that the mean differences are representative of those obtained for items varying widely in forward strength. Only if the correlation between the forward and the backward strength of individual items is assumed to be very high can these findings be generalized to the relation between degree of R-S learning and temporal changes in R-S recall.

Analysis of variance (after a Freeman-Tukey square-root transformation to remove heterogeneity of variance) yields two significant effects. The overall difference between the control condition and the combined experimental conditions is, of course, significant, F (1, 102) = 95.25. In addition, there is a significant Paradigm × Retention Interval interaction within the experimental conditions, F (1, 102) = 4.72, $p < .05$. This interaction reflects the fact that there is a decline for C-B and a rise for B-C between the immediate and the delayed test. However, when the B-C and the control condition are compared directly, the interaction with retention interval is not reliable, F = 2.04. While the lenient scores are consistently higher than the stringent ones, the pattern of results is the same and the statistical conclusions remain unchanged.

Since Ss were required to recall first-list stimuli only, intrusions from the second list were possible. In the case of C-B the intruding terms were second-list stimuli. There were 9 such cases on the immediate test and 13 on the delayed test, and these may be taken to reflect occasional failures of list differentiation. For B-C the intruding items were second-list responses. There

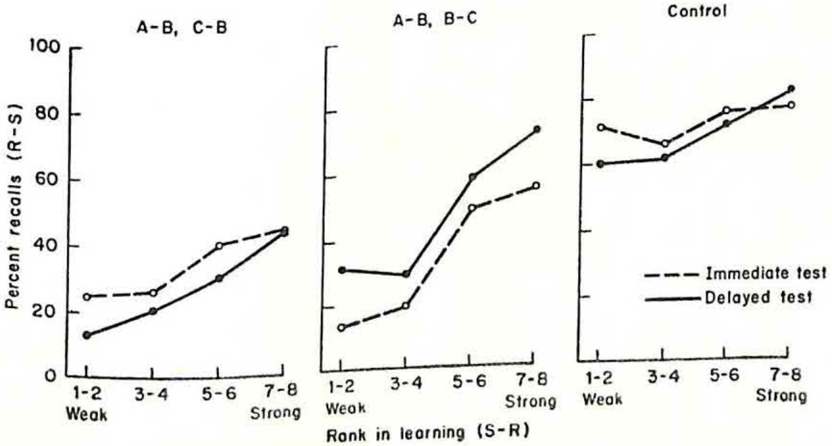


FIG. 9. Percentages of R-S recalls on the immediate and on the delayed test as a function of the S-R strength of the items in acquisition, Exp. V.

were 3 occurrences on the immediate test and 10 on the delayed test. Virtually all the intrusions for B-C were appropriately paired, and only slightly less than half were for C-B. It can be seen that the rise in first-list R-S recall occurs under the B-C paradigm in spite of some apparent loss in the ability to differentiate between the associates, one forward and one backward, of the B terms in the successive lists. The fact that all items belonged to the same class was undoubtedly conducive to the loss of differentiation. In any event, to the extent that such losses occurred, the test for absolute recovery was conservative.

The results of the statistical tests limit the conclusions that can be drawn from the observed changes in R-S recall. The significant Paradigm × Retention Interval interaction supports the argument that the processes responsible for the retroactive losses under the two paradigms are different. This inference is strengthened by the fact that RI on the immediate test turned out to be identical under the two conditions and the subsequent temporal trends were in opposite directions. Strictly speaking, it cannot be stated that R-S recall shows absolute recovery under the B-C paradigm since the interaction with retention interval fell short of significance in the direct comparison with the control condition. The evidence clearly points toward the conclusion that recovery occurred. At the same time, the limited amount of improvement, along with the outcome of the statistical tests, indicates that the recovery effect is marginal at best.

GENERAL DISCUSSION

The theoretical considerations that led to the formulation of the present experiments have been set forth in the introduction. The general discussion of the results may be conveniently cast in the form of a series of conclusions.

1. Evidence has been obtained for the recovery of first-list responses under both the A-C and the C-D paradigms. The occurrence of recovery was found to be critically dependent on the degree of first-list learning. The reversal of the conditions responsible for the retroactive losses observed immediately after the end of IL is gradual and probably only partial. Hence there appears to be, under given conditions of IL, a critical degree of OL required for the detection of recovery effects.

2. There is no evidence of greater recovery under the A-C than under the C-D paradigm. In Exp. I there was a trend in the opposite direction; in Exp. III the amounts of recovery were closely comparable. It was suggested that a theoretical model that postulates the concurrent unlearning of contextual and specific associations entails the expectation of greater recovery for A-C than for C-D. Ad hoc assumptions become necessary if the lack of a difference is to be accommodated. It could be proposed that contextual associations are extinguished more thoroughly in the A-C than in the C-D condition. Such a proposal would be inconsistent, however, with the assumption that unlearning is a function of the paradigm of transfer obtaining for a given type of associations (McGovern, 1964). With respect to contextual associations, the conditions of transfer are the same for A-C and for C-D. Alternatively, one may assume that unlearning effects for different types of associations interact in a nonadditive fashion. In that case, the analytic and predictive usefulness of the component analysis of unlearning would certainly be reduced. Finally, the possibility may be considered that extinguished specific associations recover much more slowly than contextual ones and hence contribute little to the net recovery observed over a short interval of time. There are no obvious theoretical arguments in support of this possibility. In short, it appears that the failure of the A-C condition to show greater recovery than the C-D condition poses a difficult interpretative problem for current extinction models.

3. The results may be viewed as consistent with the hypothesis that recovery represents largely or entirely the dissipation of response set interference. If the degree of response suppression varies directly with the amount of associative interference in transfer, e.g., is greater for A-C than for C-D, then the extent of recovery over a fixed period of time need not be proportional to the amount of retroactive loss observed on the immediate test. Rather the proportional gain should be greater for C-D than for A-C. In Exp. I an absolute rise in total recall scores was observed only for C-D, although the performance on the immediate test was approximately equal for the two paradigms. In Exp. III, C-D was initially superior to A-C, but the absolute amounts of gain were equal. Thus, a greater proportion of

what is initially lost does appear to recover when IL is free of heavy associative interference.

4. Manipulation of the TPI yielded results consistent with those obtained by the conventional method of measuring recovery. The effect of the temporal variable was found to be more pronounced when the TPI rather than the IL-test interval was manipulated. It also became apparent, however, that the two types of design do not provide equivalent methods for the assessment of recovery. It is true that a forward shift in the TPI reduces the length of the interval during which the processes underlying recovery can develop. At the same time, however, the delay between OL and IL may cause the first-list associations to lose strength and render them more susceptible to interference. There may also be systematic changes in the conditions of transfer during IL (Exp. II and IV). Furthermore, the conditions of list differentiation change as a function of the TPI. As the separation between OL and IL becomes greater, the original and the interfering lists acquire increasingly distinctive temporal "tags." Thus, differences in the effects of temporal manipulations obtained under the two procedures are complexly determined. With these reservations, the two designs yield results that converge in supporting the empirical fact of first-list recovery.

5. Systematic shifts in the order of output of the responses from the successive lists accompany recovery, reflecting a reduction of the bias in favor of the more recently learned material. These changes are fully consistent with the hypothesis of response set interference, but could in principle be attributed to gains in associative strength as well. The nature of the temporal changes in the degree of stimulus-determined clustering points to the effectiveness of nonassociative factors in governing the order of output: the degree of clustering declines as recall increases. It has already been suggested that the introduction of an IL-test interval may enhance *S*'s disposition to scan available responses and to match them to the stimulus terms. The more general point warranting consideration is that the delay serves to weaken the constraints that the physical and temporal arrangements of the acquisition procedure impose on *S*'s mode of performance. Such processes as repeated response scanning and associative matching are increasingly likely to come into play as the delay is lengthened. There develops, as it were, a gradual shift from strictly associative responding to retrieval and matching of responses. Thus, recovery may reflect not only the dissipation of response set interference, but also a beneficial change in the mode of recalling.

6. The temporal changes in R-S recall bear out the view that functionally different mechanisms are responsible for the losses produced by IL under the C-B and B-C paradigms. The results are consistent with the hypothesis that in the case of B-C, a tendency established during acquisition to suppress stimulus terms gradually dissipates, whereas the losses for C-B reflect a process of reproductive inhibition localized in the test stage. There was only marginal evidence, however, for absolute recovery in the B-C condition. To put this finding into perspective, it is well to bear in mind that the backward associations established during OL in the present experiment were weak relative to the forward ones, as is evidenced by the level of recall of the control group. Hence backward associations were more vulnerable than forward ones to RI and benefited less from the gradual dissipation of interference over time.

7. The comparisons of recovery effects obtained under various paradigms bring into focus a general problem of inference that is likely to arise when a difference, or more particularly a lack of difference, between amounts of interference is to be interpreted. When different manipulations produce equal amounts of interference, it is clearly hazardous to conclude that the underlying mechanisms are the same in each case. Conditions yielding identical amounts of retention loss on an immediate test may diverge over time (Exp. I and V). The analysis of the course of recovery may thus provide a sensitive method for assessing the mechanisms of interference coming into operation under different conditions of transfer.

REFERENCES

Abra, J. C. Time changes in the strength of forward and backward associations. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 640–645.

Abra, J. C. Time changes in the strength of extinguished context and specific associations. *Journal of Experimental Psychology*, 1968, 77, 684-686.

Archer, E. J., & Underwood, B. J. Retroactive inhibition of verbal associations as a multiple function of temporal point of interpolation and degree of interpolated learning. *Journal of Experimental Psychology*, 1951, 42, 283–290.

Cheung, C. C., & Goulet, L. R. Retroactive inhibition of R-S associations in the A-B, B-C

C-B paradigms. *Journal of Experimental Psychology,* 1968, **76**, 321–322.

Ellington, N. R., & Kausler, D. H. "Fate" of List 1 associations in transfer theory. *Journal of Experimental Psychology,* 1965, **69**, 207–208.

Greenbloom, R., & Kimble, G. A. Extinction of R-S associations and performance on recall trials without immediate feedback. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 341–347.

Guilford, J. P. *Fundamental statistics in psychology and education.* (2nd ed.) New York: McGraw-Hill, 1950.

Houston, J. P. Unlearning of specific associations in the A-B, C-D paradigm. *Journal of Experimental Psychology,* 1967, **74**, 254–258. (a)

Houston, J. P. Proactive inhibition and competition at recall. *Journal of Experimental Psychology,* 1967, **75**, 118–121. (b)

Houston, J. P. Retroactive inhibition and point of interpolation. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 84–88. (c)

Keppel, G. A reconsideration of the extinction-recovery theory. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 476–486.

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Keppel, G., & Schwartz, B. Response availability and conditions of interpolated practice. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 489–493.

Keppel, G., & Underwood, B. J. Retroactive inhibition of R-S associations. *Journal of Experimental Psychology,* 1962, **64**, 400–404.

Koppenaal, R. J. Time changes in strengths of A-B, A-C lists; spontaneous recovery? *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 310–319.

McGovern, J. B. Extinction of associations in four transfer paradigms. *Psychological Monographs,* 1964, **78**(16, Whole No. 593).

Newton, J. M., & Wickens, D. D. Retroactive inhibition as a function of the temporal position of interpolated learning. *Journal of Experimental Psychology,* 1956, **51**, 149–154.

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

Postman, L. Unlearning under conditions of successive interpolation. *Journal of Experimental Psychology,* 1965, **70**, 237–245.

Postman, L., & Stark, K. Role of response availability in transfer and interference. *Journal of Experimental Psychology,* 1969, **79**, 168–177.

Postman, L., Stark, K., & Fraser, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 672–694.

Saltz, E., & Hamilton, H. Spontaneous recovery of List 1 responses in the A-B, A′-C paradigm. *Journal of Experimental Psychology,* 1967, **75**, 267–273.

Schwartz, M. Effect of stimulus class on transfer and RI in the A-B, C-B paradigm. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 189–195.

Twedt, H. M., & Underwood, B. J. Mixed vs. unmixed lists in transfer studies. *Journal of Experimental Psychology,* 1959, **48**, 111–116.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs,* 1945, **59**(3, Whole No. 273).

Underwood, B. J., & Ekstrand, B. R. Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology,* 1967, **74**, 574–580.

Wollen, K. A., & Gallup, G. G., Jr. R-S recall as a function of opportunity for intratrial rehearsal. *Psychonomic Science,* 1965, **3**, 157–158.

Wollen, K. A., & Gallup, G. G., Jr. R-S recall as a function of presence or absence of successive pair repetitions in S-R learning. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 77–80.

(Received May 12, 1969)

334

24 · 8 · 71

bele

Journal of
Experimental Psychology

VOL. 82, NO. 2 NOVEMBER 1969

SUBJECT PREFERENCES AND THE NATURE OF INFORMATION STORED IN SHORT-TERM MEMORY[1]

KENNETH R. LAUGHERY[2] AND JAMES C. FELL[3]

Department of Industrial Engineering, State University of New York at Buffalo

A two-part experiment was carried out to explore performance on a memory-span task as a function of the presentation mode that *S*s prefer. The first part of the experiment used a memory-span procedure involving eight-letter sequences where the presentation mode (visual or auditory) was randomly mixed. In the second task, *S*s were exposed to auditory and visual five-letter sequences simultaneously and could record either the auditory or visual sequence—establishing a preference. The compatibility between mode preference in the second task and relative performance on visual and auditory items in the first task showed that *S*s perform better on auditory items regardless of their mode preference. This result is interpreted as support for the notion that short-term memory is basically an auditory storage system.

A great deal of experimental and theoretical work has been reported in the past several years on the nature of information stored in short-term memory (STM). Several theories (Atkinson & Shiffrin, 1968; Laughery, 1969; Sperling, 1967) have proposed that STM is primarily an auditory storage system; i.e., it consists of an encoding of the auditory characteristics of the information.

There are several lines of evidence that have led to the conclusion that STM is basically an auditory representation. Analyses of intrusion errors in memory-span or modified-span studies (e.g., Wickelgren, 1965a) have shown that incorrect items tend to have auditory characteristics in common with the correct items for which they were substituted. Such analyses have failed to show similarities based on visual dimensions. A second series of experiments, also employing a memory-span procedure, has explored the effects of interitem similarity on performance. Experiments in which the auditory similarity between items within a sequence was varied showed that performance under conditions where items were auditorily distinct was superior to conditions where items were auditorily similar (Conrad & Hull, 1964; Wickelgren, 1965b). Other experiments (Baddeley, 1966; Cimbalo & Laughery, 1967) have shown little or no effect of visual similarity between items on performance. Murdock (1968) reported results representative of a third

---

[1] This research was supported by Research Grant MH-11595 from the National Institute of Mental Health, United States Public Health Service.

[2] Requests for reprints should be sent to Kenneth R. Laughery, Department of Industrial Engineering, State University of New York at Buffalo, Buffalo, New York 14214.

[3] Now at Cornell Aeronautical Laboratory, Inc., Buffalo, New York.

© 1969 by the American Psychological Association, Inc.

line of evidence related to the issue. Using a probe technique, Murdock compared performance as a function of the mode, auditory or visual, in which the information was presented. The results showed that performance is superior for the auditory presentation mode and suggest that the modality effect represents differences in storage as opposed to retrieval. Murdock's results can be viewed as supporting the suggestion that STM is basically an auditory storage, since the auditory presentation is compatible with the information being stored.

One aspect of the issue as to the nature of information stored in STM that has not been explored to date concerns how *S*s *prefer* to process or deal with information in STM. If *S* is viewed as an active processor with control over what gets into STM, the results of the aforementioned experiments may be due to the fact that most *S*s prefer to auditorily encode the information. If one can control or identify the information that *S*s process or store, and if some *S*s process or encode the information visually, it may be possible to show that the results of the aforementioned experiments reflect a processing strategy as opposed to a basic characteristic or property of STM.

One line of research that has prompted this approach to the issue is the work on visual imagery reported by Bugelski, Kidd, and Segmen (1968) and Paivio (1968). The results of these studies indicate that performance in a memory task is greatly improved when *S*s are instructed to visually encode (form images of) the information.

The present study is an attempt to determine if *S*s have preferences for processing information in a particular way, if *S*s differ in these preferences, and if performance is a function of these preferences.

METHOD

Subjects.—Sixty-seven students enrolled in an introductory psychology course at the State University of New York at Buffalo participated in the experiment to fulfill a course requirement.

Design.—There were two distinct parts to the experiment. The first phase consisted of a single factorial design where the variable manipulated was presentation rate. Five levels of rate were used (.5, .75, 1.0, 2.0, and 3.0 sec/item). The number of *S*s per group ranged from 7 to 19 (i.e., rate was manipulated as a between-*S*s variable).

In the second phase of the experiment a similar design was used, although a somewhat different task was involved (described subsequently). The same five levels of presentation rate represented the single factor in this phase.

Procedure.—The first phase of the experiment consisted of a memory-span procedure in which *S*s were exposed to 15 eight-letter sequences. The presentation mode, visual or auditory, was mixed within each sequence. In other words, within a sequence some of the items were presented visually and others auditorily. At the end of each sequence *S*s were required to write down as many of the items as they could remember. The *S*s recorded their responses on answer sheets containing eight spaces, and instructions were given to guess when necessary.

The second phase of the study was a mode-preference test and was administered immediately after Phase 1 was completed. In this phase, *S*s were exposed to 15 sequences of five letters presented visually and five letters presented auditorily. The auditory and visual sequences were presented simultaneously and consisted of the same five letters but in different orders. The rate at which the sequences were presented was the same as *S* had received in Phase 1 of the experiment. The *S*s were told that the visual and auditory sequences would contain the same set of five letters but in different random orders. They were further instructed that they could record the items at the end of the presentation in the order that they were visually presented, in the order that they were auditorily presented, or in a mixed order. In other words, *S* was free to record the five letters in any way he chose.

At the end of Phase 1, *S*s were asked to write down any strategies that they used to remember the letters. Following Phase 2, *S*s were asked to note the order in which they preferred to remember the sequences and why.

Apparatus.—The visual items were generated on an in-line digital display unit (Binaview) where the item was a lighted letter on a dark background. A closed-circuit television system was used for presentation. The letters appeared on a 19-in. television set placed in a university classroom. The letters were approximately 9 in. high on the television screen. The auditory items were recorded on a Roberts 1057 tape recorder. A speaker was connected to the tape recorder to provide a better quality of auditory presentation.

The speaker was placed to the left of the television set on a table in the front of the room. Synchronization between the auditory and visual items was possible through the use of a Photo Sync on the tape recorder. Small pieces of sensing tape placed at timed intervals on the recording tapes set off the visual presentations.

Materials.—Fifteen sequences of eight nonrepeating, randomly drawn letters were used. The auditory and visual item presentations were randomized within a sequence, with the restriction that neither mode appeared more than five times (or less than three times) within a sequence. Each of the five groups of *S*s received the same 15 sequences. Each *S* was given answer sheets which contained columns indicating where to record each sequence at the time of recall.

For the mode-preference study, 15 sequences of five nonrepeating, random consonants were used. For each sequence, one order of the five consonants was presented visually while *another order* of the *same* five consonants was simultaneously presented auditorily. A stipulation was that at each point in time a *different* consonant appeared visually and auditorily. Answer sheets indicating where to record the sequences were given to each *S*. The same sequences and orders of presentation were used in each of the five groups; only the rates were different.

RESULTS

In order to minimize practice effects, only the last 10 sequences of each *S*'s data in the memory-span task were analyzed. The criterion was the number of correct items in the proper position. An analysis of variance indicated a significant presentation rate effect, F (4, 62) $= 7.09$, $p < .005$, where the means were 3.4, 4.4, 3.9, 4.8, and 6.0 for the .5, .75, 1.0, 2.0, and 3.0 rates, respectively.

A more detailed analysis was carried out in which performance on the visually presented items was compared to performance on the items presented auditorily. Since the number of auditory and visual presentations for the last 10 trials were not completely balanced (46 auditory items vs. 34 visual items), percentage correct for each mode was used as a criterion. The results are plotted in Fig. 1. An analysis of variance indicated that presentation mode, F (1, 62) $= 46.7$, $p < .001$, and presentation rate, F (4, 62) $= 5.2$, $p < .005$, were significant. The Mode × Rate interaction was also significant, F (4, 62) $= 4.69$, $p < .005$, with the difference between auditory and visual decreasing at the slower rates.

The results of the second phase of the study, the mode-preference task, were analyzed to determine which, if either, mode *S*s preferred to process. The last 10 sequences were analyzed for each *S* and the percentage of sequences recorded in the auditory order, visual order, or in a mixed order were determined. The results are shown in Fig. 2. Clearly, the preference is for the auditory mode at the faster rates, with the difference decreasing at the slower rates until at a 3.0-sec. presentation there is no clear-cut general preference. A performance analysis based on the number of correct items indicated no differences due to rate. This was expected, however, since most *S*s can recall five items even at the fast presentation rates (the mean number correct for the different rate conditions ranged from 4.73 to 4.84, indicating a clear-cut ceiling effect).

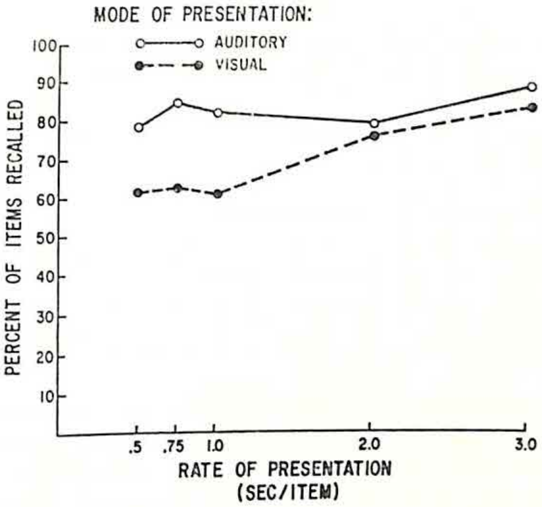


FIG. 1. Percentage of items correct for each mode in the memory-span task.

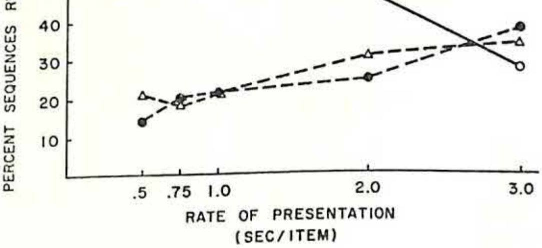


FIG. 2. Percentage of sequences recorded on the basis of presentation mode in the mode-preference task.

TABLE 1

MODE-PREFERENCE ANALYSES: NUMBER OF *S*s IN EACH PREFERENCE CATEGORY

Presentation rate	Classification based on protocols			Classification based on recall order		
	Aud	Vis	Mix	Aud	Vis	Mix
.5	14	2	3	15	2	2
.75	10	3	1	9	4	1
1.0	5	0	2	5	1	1
2.0	6	3	1	5	2	3
3.0	5	6	6	4	6	7

Based on the order in which the five-item sequences in the mode-preference task were recorded, each *S* was put into an auditory, visual, or mixed preference category. The rule for assigning an *S* to one of these categories was that if 6 or more of the 10 sequences were recorded in a particular mode, *S* was assigned to that category. If an *S* recorded half of the sequences in each mode or if items were mixed within sequences, *S* was placed in the mixed category.

The protocol obtained at the end of the mode-preference task was examined and also served as a basis for assigning *S*s to the different preference categories. The results of these classification procedures are shown in Table 1. It can be seen that there was a great deal of consistency between the assignments made on the basis of the protocols and the assignments based on performance in the task.

Finally, an analysis was carried out to explore preference-performance compatibility, i.e., how well *S*s recall the visual vs. auditory items in the memory-span task as a function of their preferred mode as identified in the mode-preference task. For example, if an *S* preferred the auditory mode and also recalled more auditory than visual letters by a factor of 10%, he would be considered in the aud-aud category and his preference-performance would be compatible. If the performance in the two modes did not differ by more than 10%, the performance was classified as "even." The compatibility analysis is shown in Table 2. Of the 53 *S*s who could clearly be categorized as auditory or visual with respect to preference, the preference-performance relationship was compatible for 33, incompatible for 10, and unrelated for another 10. From the table, only 2 *S*s preferred the auditory mode, but performed better on the visual mode; while 8 *S*s preferred the visual mode, but performed better on the auditory mode.

TABLE 2

PREFERENCE-PERFORMANCE COMPATIBILITY ANALYSES: NUMBER OF *S*s WITH VARIOUS PREFERENCE-PERFORMANCE RELATIONSHIPS FOR EACH PRESENTATION RATE CONDITION

Preference-performance	Rate					
	.5	.75	1.0	2.0	3.0	Total
Aud-Aud	14	9	4	2	2	31
Aud-Vis	1	0	0	1	0	2
Aud-Even	0	0	1	2	2	5
Vis-Aud	1	4	0	1	2	8
Vis-Vis	1	0	0	0	1	2
Vis-Even	0	0	1	1	3	5
Mix-Aud	1	1	1	1	3	7
Mix-Vis	0	0	0	0	0	0
Mix-Even	1	0	0	2	4	7

To test the hypothesis that one thing remembered about an item in an STM task is the mode in which it was presented, an additional analysis was carried out on mode intrusion errors. Mode intrusion errors were errors made by *S*s when they substituted a letter of the same mode for the correct letter. Therefore, an auditory intrusion occurred when an *S* substituted another letter he *heard* in the sequence for the correct auditory letter. Similarly, a visual intrusion occurred when an *S* substituted another letter he *saw* in the sequence for the correct visual letter. These auditory and visual intrusions were collapsed over presentation rate and are referred to as "mode intrusion errors." A percentage of mode intrusion errors over the total number of order errors for each individual *S*, regardless of mode, was calculated. A *t* test was carried out to determine if the occurrence of mode intrusion errors was greater than chance. The result was significant, $t = 2.64$, $p < .01$, indicating that presentation mode is a relevant attribute of the information stored in STM.

DISCUSSION

The results of the present study are consistent with previous findings that emphasize the importance of auditory encoding in STM. Clearly, *S*s tend to prefer to process information in an auditory mode, at least at the faster presentation rates, and they perform better on items that are presented auditorily as compared to those presented visually.

A possible explanation for the significant interactions between preference and rate in the mode-preference study and between mode of presentation and rate in the memory-span study is based on a translation process that may occur when the information is visually presented. Since at the slower presentation rates there is ample opportunity for *S* to translate the visual stimulus to an internal auditory representation which is stored in STM, and since this translation probably takes place very quickly and leaves sufficient time for other processes such as rehearsal to be carried out, performance will not be greatly affected by the presentation mode. At faster presentation rates, however, the translation time becomes more critical in the overall performance. Thus, at fast rates *S*s will probably prefer to receive the information in the same mode as it is processed, since this will result in a higher level of performance.

Of greater interest in the present study, however, is the relative performance of those *S*s who "prefer" to process the information in an auditory mode as opposed to a visual mode. Thirty-one *S*s preferred and did better on the auditory mode, while only 2 *S*s preferred and performed better on the visual mode. Some other comparisons are even more interesting. Of the 52 *S*s who could be classified as having either auditory or mixed preferences, only 2 performed better on the visual items. On the other hand, of the 29 *S*s who could be classified as having visual or mixed preferences, 15 performed better on the auditory items. Looking only at those whose preferences were mixed, 7 performed better on the auditory items, while none performed better on the visual items. Finally, only 2 *S*s preferred the auditory and performed better on the visual items, whereas 8 preferred the visual, but performed better on the auditory items.

All of these observations indicate that mode preferences which *S*s have for processing alphanumeric information do not play a crucial role in STM. Rather, it appears that the storage of auditory information is a basic characteristic or property of STM.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968.

Baddeley, A. D. Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology,* 1966, **18**, 362–365.

Bugelski, B. R., Kidd, E., & Segmen, J. Image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology,* 1968, **76**, 69–73.

Cimbalo, R. S., & Laughery, K. R. Short-term memory: Effects of auditory and visual similarity. *Psychonomic Science,* 1967, **8**, 57–58.

Conrad, R., & Hull, A. J. Information, acoustic confusion and memory span. *British Journal of Psychology,* 1964, **55**, 429–432.

Laughery, K. R. Computer simulation of short-term memory: A component-decay model. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 3. New York: Academic Press, 1969.

Murdock, B. B., Jr. Modality effects in short-term memory: Storage or retrieval? *Journal of Experimental Psychology,* 1968, **77**, 79–86.

Paivio, A. Effects of imagery instructions and concreteness of memory pegs in a mnemonic system. *Proceedings of the 76th Annual Convention of the American Psychological Association,* 1968, **3**, 77–78.

Sperling, G. Successive approximations to a model for short-term memory. *Acta Psychologica,* 1967, **27**, 285–292.

Wickelgren, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology,* 1965, **70**, 102–108. (a)

Wickelgren, W. A. Short-term memory for phonemically similar lists. *American Journal of Psychology,* 1965, **78**, 567–574. (b)

(Received March 10, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 198–204

HYDRO-HYPODYNAMIC SENSORY ISOLATION EFFECTS ON CONCEPT IDENTIFICATION[1]

VLADIMIR PISHKIN[2] AND JAY T. SHURLEY[3]

Veterans Administration Hospital, Oklahoma City

One hundred and eighty male medical students participated in a study designed to test the effects of hydro-hypodynamic environment (HHDE) on auditory concept identification (CI). The major findings were: (*a*) There was no difference in CI performance between HHDE conditions with immediate onset of CI problem and standard laboratory procedure; (*b*) performance improved after 30 min. in HHDE as compared to standard laboratory and control HHDE conditions; (*c*) errors significantly increased after 60 min. with some improvement after 120 min. in HHDE; (*d*) significant interaction effects were found between complexity of CI task and duration in HHDE in both CI errors and time estimation scores; and (*e*) *S*s reliably underestimated time spent in HHDE.

The present experiment was conceived to investigate the auditory concept identification (CI) performance of human *S*s in a profound sensory isolation (SI) environment as a function of information complexity and duration in SI.

There is ample evidence that for many *S*s SI is a stressful procedure resulting in restlessness, inability to concentrate, disorientation in time, and decline in intellectual performance (Scott, Bexton, Heron, & Doane, 1959). Pishkin and Shurley (1966a) demonstrated reduction in operant conditioning rates of rats as a function of SI. Moreover, Rossi, Furhman, and Soloman (1967), using electro-oculographic and electroencephalographic recordings, have shown that SI does not disrupt relationship between levels of arousal and thought processes, although mental experiences may be misinterpreted by *S*s in the absence of usual cues. Although general evidence indicates that one should expect inhibiting effects of SI on cognitive performance, there is also evidence that SI may facilitate rote learning as the confinement continues up to 48 hr. (Vernon, 1963, p. 86).

There is conflicting evidence concerning cognitive and perceptual functions in *S*s exposed to environments in which, to some degree, there is reduction of sensory stimulation or general lack of information input. Notably, recent experiments dealing with reduction of usual environmental cues were aimed at understanding the influences of SI in relation to a variety of psychological phenomena such as arousal levels (Rossi et al., 1967), verbal learning (Vernon, 1963), thought disorganization in "acute sensory deprivation" (Ziskind, 1965), and shift from conceptual to perceptual mode thought processes (Shurley, 1960; 1966).

The present state of research in sensory and perceptual deprivation and isolation indicates that, in general, performance on cognitive tasks is either impaired, unaffected, or facilitated, depending on the nature of the task. Equally important is a notion of some investigators that SI belongs in the class of drive-enhancing arousing phenomena (Lilly, 1956; Suedfeld, Glucksberg, & Vernon, 1967). Presumably, general drive level increases with SI duration. Within the Yerkes-Dodson (1908) paradigm, it would be expected that SI would facilitate or impair task performance nonmonotonically depending on the drive level. Suedfeld et al. found the effects of incentive and SI to be similar in a problem-solving situation. Performance improved as function both of increased incentive and of SI; thus, they also concluded that SI can be functionally related to traditional drive functions. It is important to note also that simple rote learning and recall tasks are facilitated by SI, while performances in relatively more complex tasks are impaired with SI (Suedfeld, 1964). Accordingly, it is a plausible expectation that performance of relatively simple cognitive problems would be less likely to be affected adversely by SI than more complex problems.

---

[1] The authors are indebted to Lyle E. Bourne, Jr., for his thoughtful suggestions in reviewing of the manuscript. Thanks are due to Lewis M. Jones, Aaron Wolfgang, and W. Vail Williams for their assistance in collecting the data.

[2] Requests for reprints should be sent to Vladimir Pishkin, Veterans Administration Hospital, 921 Northeast 13th Street, Oklahoma City, Oklahoma 73104.

[3] The authors are also affiliated with the University of Oklahoma School of Medicine.

The present investigation was conceived primarily to study the effects of profound SI in the hydro-hypodynamic environment (HHDE; Shurley, 1961) on the identification of auditory concepts (CI). Task complexity and duration of SI were the main independent variables. Secondarily, the investigation was designed to study the effects of SI on subjective estimation of elapsed time as a function of task complexity.

METHOD

Subjects.—One hundred and eighty male medical students from the University of Oklahoma School of Medicine volunteered to serve as paid *S*s ($3.00/hr) and were randomly assigned in groups of 3 *S*s to the 60 experimental conditions.

Design.—This was a 5 × 4 × 3 factorial design with 5 duration conditions: 0, control, 30, 60, or 120 min. in SI; 4 levels of complexity: 0, 1, 2, or 3 irrelevant dimensions; and 3 problems: duration, frequency, and number were used as the relevant dimensions in the three problems. The different problem types were introduced only to minimize spread of information among *S*s, since it was found in a previous study that they were not a significant source of variance (Pishkin & Shurley, 1965).

Task and apparatus.—The 0-hr. SI group (0-SI), consisting of 36 *S*s, reported to the laboratory and performed on the auditory concept identification (CI) task in the standard laboratory setting (Pishkin & Shurley, 1965). The control group, consisting of 36 *S*s, reported directly to the SI tank and followed exactly the same procedure (described in "SI procedure" section) as the HHDE *S*s with the exception that the CI task commenced immediately after *S*s were isolated (i.e., SI chamber door closed and lights turned out). The remaining 108 *S*s, all of whom were exposed to SI condition, performed on a CI task in the HHDE after 30 (30-SI), 60 (60-SI), or 120 min. (120-SI) of continuous SI exposure.

The 0-SI group performed on a CI task in the following manner. The *S*s were seated in a comfortable reclining chair located in a sound-attenuated room and wore headphones throughout the experiment. They were given the following CI instructions by the male *E*:

> You will hear a series of tones through these earphones. Your task is to classify these tones into two categories. When the tone signal stops, this light will go on, indicating to you that you are to respond by saying either "X" or "O." These tones may vary in different ways. They may be long or short, high or low in frequency, one or two in number. Only one of these ways will help you to categorize correctly. If you expect a particular signal to belong in the X category, when this light goes on, you will say "X." After you respond, either letter X or O will appear on this screen. If you said "X," and the letter X appears, you will be correct. If you said "O," and the letter O appears, you will also be correct. However, if you said "X," and the letter O appears, you will be wrong. The same holds true if you said "O," and the letter X appears. Your job is to be correct as often as possible. When this light goes on you are to respond. For a time you will be guessing. However, your guesses may turn out to be correct. Do you have any questions?

Instructions were elaborated to those *S*s who could not understand the task, meaning of the feedback lights, etc. The same instructions were given in the SI environment, with the exception that *S*s were told that they would hear a warning tone (5 sec.) before CI procedure was initiated. They were not told how long they would be in SI before the CI task would start. All *S*s were *naive*, and they had not previously been exposed to any CI studies, visual or auditory, nor to SI or social isolation.

The relevant dimensions for the three problems were duration of signal (1 or 3 sec.); frequency (1,000 or 3,000 cps); and number (1 or 2 signals for a duration of 1 sec. with a 1-sec. interval between them). Number was irrelevant only with two or three irrelevant-dimension problems. When a dimension was neither relevant nor irrelevant, it was always presented at one of its two levels. The levels were 1 sec. in duration on both earphones for laterality, 1,000 cps in frequency, 80 db. (SPL) in intensity (this dimension was not varied), and 1 signal in number. When laterality was used as irrelevant dimension, tones were presented in either the right or the left earphone; this dimension was used only in three irrelevant-dimension problems. An auxiliary control circuit was built into the Ensco Model PC-1 Phantom Sound Unit, which consisted of a time delay relay controlled by a thyratron. The two resistance-capacitance networks controlled the thyratron in order to provide either 1- or 3-sec. duration pulses. The sound unit was also connected to a control box whereby *E* could select either a 1,000-cps or 3,000-cps tone.

The *S*s were presented with one of the three problems where either duration, frequency, or number dimension was relevant. Within each problem, *S*s were assigned to four equal groups whose problem contained either 0, 1, 2, or 3 irrelevant dimensions. Thus the experiment contained the following 12 combinations of dimensions: (*a*) duration relevant, nothing irrelevant; (*b*) frequency relevant, nothing irrelevant; (*c*) number relevant, nothing irrelevant; (*d*) duration relevant, frequency irrelevant; (*e*) frequency relevant, duration irrelevant; (*f*) number relevant, duration irrelevant; (*g*) duration relevant, frequency and number irrelevant; (*h*) frequency relevant, duration and number irrelevant; (*i*) number relevant, duration and frequency irrelevant; (*j*) duration relevant, frequency, number, and laterality irrelevant; (*k*) frequency relevant, duration, number, and laterality irrelevant; and (*l*) number relevant, duration, frequency, and laterality irrelevant.

The sequence of tone presentations was random with restrictions that identical stimuli never occurred on successive trials; X and O responses were balanced. In the 0-SI condition, *E* sat and manipulated the controls from a position behind and out of sight of *S*. The feedback box (6 × 5 × 4 in.) was positioned on the ceiling above and in front of *S*. The letter X or O was illuminated by *E* following *S*'s responses. The task was self-paced, and on termination of the signal, a ready light on the feedback box was lighted for 1 sec., indicating to *S* to respond to the preceding stimulus. The criterion to solution was set at 16 consecutive correct responses. In the SI situation, *E* and controls were in the anteroom, and the feedback box was attached, as in the laboratory, to the ceiling above *S* in the tank.

SI procedure.—For SI conditions, *S*s reported to the laboratory and were asked to urinate before proceeding into the water in their swimming shorts. The HHDE was designed to achieve maximal diminution of all possible sensory input. It consisted of two rooms: the "SI chamber" was a lightproof, soundproof, 12 × 20 × 9 ft. room with an 8 × 8 ft. cylindrical water tank located in the center. The slowly flowing water was maintained at a constant temperature of 94° F. The second room was an antechamber where *E* manipulated the CI task controls. The antechamber was separated from the SI chamber by a heavy, insulated (cold-storage type) door and a soundproof wall. Each *S* wore a bathing suit and was positioned on soft, buoyant plastic foam strips under hips and neck to remain motionless at neutral buoyancy, face only above water. He was under instruction to inhibit body movements, consistent with comfortable positioning, and to avoid painful discomfort. Waterproof earphones were placed on *S*, and the standard CI instructions were given. The *S* was then told that *E* would leave the room and the lights would be turned out. Then, after a certain period of time, *S* would hear a 5-sec. tone, at which time he was to say aloud that he was "ready." A microphone was suspended directly above the tank, and *S* was told that *E* would be in the adjacent antechamber throughout the experiment and would be able constantly to hear *S*, but would not communicate with him. At the end of the CI procedure, lights were turned on and *S* was asked to estimate his total time in the tank. A more detailed description of the procedure in the SI tank may be found elsewhere (Shurley, 1960).

RESULTS

The two major dependent variables analyzed in the present study were (*a*) number of errors on the CI task and (*b*) the discrepancy between the actual and the perceived time in SI.

Concept identification.—Mean errors by the four SI-duration groups as functions of four levels of complexity are represented in Fig. 1. It will be noted that there is a general tendency for errors to decrease from 0-SI and control to 30 min. in SI, followed by upward error trend from 30 to 60 min. and again a downward trend from 60 to 120 min. in SI. Tukey B test (Winer, 1962, p. 87) reveals, as does Fig. 1, that there was a significant improvement in performance from 0-SI and control conditions to the 30-min. condition and an increase in errors in 60-SI condition; in all comparisons (Tukey B) the 30-min. SI condition differed significantly (less errors) from the 0 ($p < .01$), control ($p < .05$), 60- ($p < .01$), and 120-min. ($p < .05$) groups. There were no significant differences between any other pairs of duration conditions. The error data were subjected to an analysis of variance, which confirmed the reliability of trends described in Fig. 1. A summary of the analysis appears in Table 1.

Only the main effects of complexity and duration were found to be significant. Subsequent orthogonal polynomial analyses were performed on the functions of complexity and duration, where the linear component of complexity reached a reliable level. This agrees with previous work on effects of this variable in visual (Archer, Bourne, & Brown, 1955; Pishkin, 1960) and auditory (e.g., Bulgarella & Archer, 1962; Pishkin & Shurley, 1965) problems. The quartic component of the duration main effect reached an acceptable level of significance, as seen in Fig. 1.

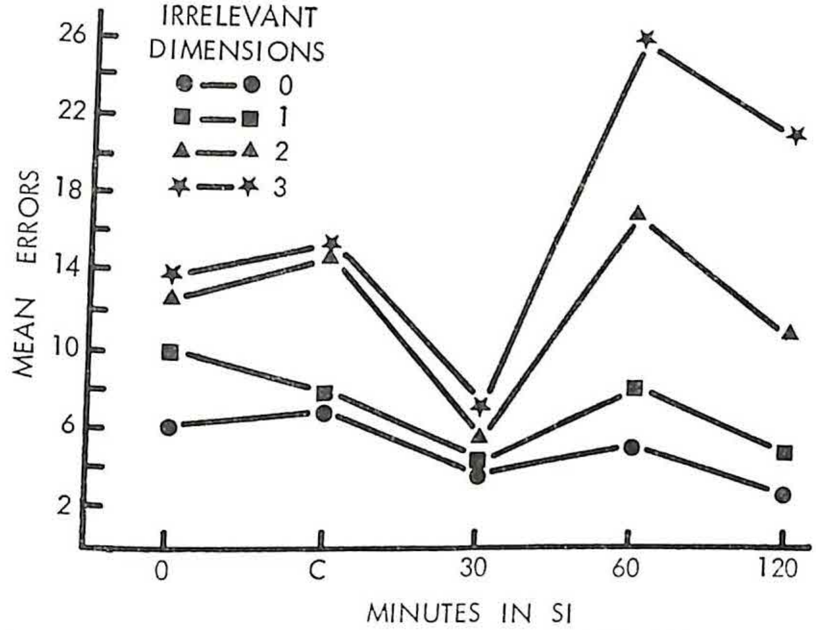


FIG. 1. Mean errors to solution by the 0-SI, control, 30-SI, 60-SI, and 120-SI groups exposed to 0, 1, 2, or 3 irrelevant dimensions.

This significant component reflects fluctuation in errors as indicated by the analysis of variance.

The significant Duration × Complexity interaction indicated increases in number of irrelevant dimensions had a greater retarding influence on CI performance, especially in the 60-min. SI duration conditions, as compared to the other conditions. In addition, a partitioned analysis of variance was performed on the 0 and the control conditions, indicating that there was no significant Duration × Complexity interaction in these conditions, $F\ (3, 120) = .41, p > .05$.

TABLE 1

ANALYSES OF VARIANCE ON CI ERRORS AND TIME DIFFERENCE SCORES

Source	CI errors			Time estimation		
	df	*MS*	*F*	*df*	*MS*	*F*
Complexity (C)	3	261.52	12.25**	3	548.56	4.65**
Linear	1	781.83	36.62**	1	1,622.44	13.75**
Quadratic	1	1.78	<1	1	23.05	<1
Cubic	1	.95	<1	1	.18	<1
Duration (D)	4	173.79	8.14**	2	969.17	8.21**
Linear	1	5.16	<1	1	1,875.18	15.89**
Quadratic	1	4.48	<1	1	63.14	<1
Cubic	1	51.93	2.43			
Quartic	1	633.57	29.68**			
Problems (P)	2	14.85	<1	2	83.41	<1
C × D	12	102.69	4.81*	6	486.97	4.13*
C × P	6	14.52	<1	6	53.08	<1
P × D	8	17.04	<1	4	71.36	<1
C × P × D	24	41.16	1.93	12	186.46	1.58
Error	120	31.25		72	118.01	
Overall time				1	(569,870.29)	48.29**

Note: To determine overall discrepancies between estimated and actual time, F test was based on the ratio between the analysis of variance correction term, $(\Sigma X)^2/N$, and the error term.
* $p < .01$.
** $p < .001$.

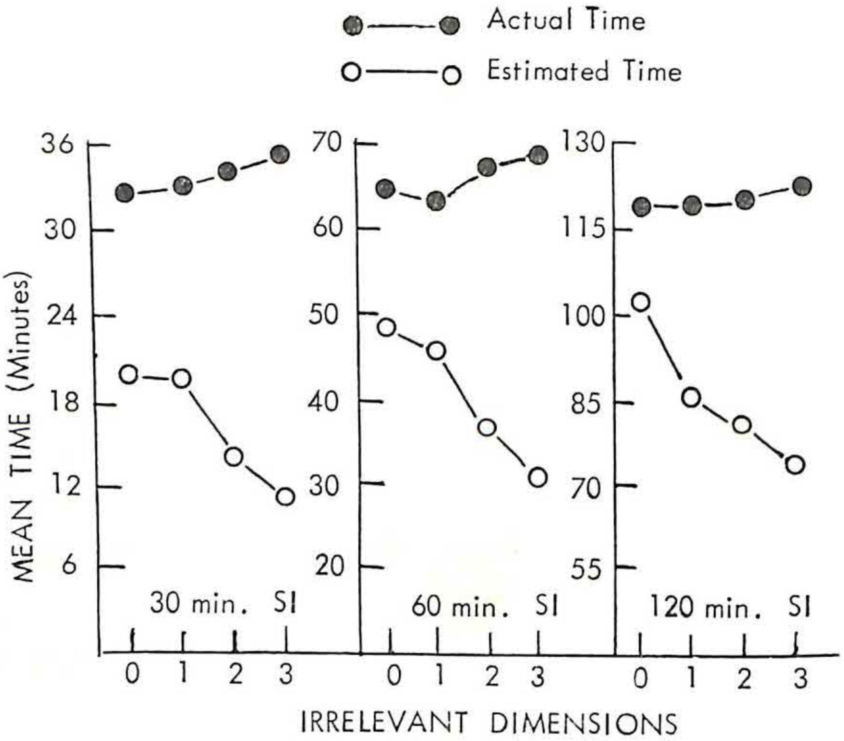


FIG. 2. Mean actual and estimated time by the four complexity groups exposed to three duration conditions in SI.

Time estimation.—Since in the control group the CI task was introduced immediately following instructions and there was no SI per se, time estimation scores were not included in the analysis of variance. The mean estimated time was 5 min. 18 sec. and the mean actual time was 5 min. 41 sec., with all complexity groups pooled; none of the discrepancies within any of the complexity groups reached significance (all t's $> .30$). Thus, it may be concluded that for such short durations, time estimates in HHDE concomitant with CI are accurate. Figure 2 represents differences between mean actual and estimated times for the three SI duration conditions (30, 60, and 120 min.), which were elaborated into four levels of CI problem complexity. For the purposes of analysis of variance on the time estimation difference scores, a constant (20 min.) was added to each actual time score, since 6 out of 108 *S*s overestimated time in the tank (by a maximum of 14 min.). The analysis of variance of these difference scores, again as functions of problem complexity, duration, and problems, is summarized in Table 1. As it is apparent from Fig. 2, the main effects of complexity and duration were significant and both functions were linear. The mean actual time varied slightly across complexity conditions since *S*s performed on CI in the tank and remained there until solution was reached. Not only was there significantly greater discrepancy between perceived and actual times, as duration in SI and number of irrelevant CI dimensions increased, but the significant Duration × Complexity interaction indicates that perceived time in SI is underestimated to a greater degree following complex tasks as the duration in SI increases. The "overall time" F test revealed that there was a significant overall discrepancy between actual and estimated times, with an overwhelming tendency of *S*s to underestimate.

DISCUSSION

Using rats in two earlier studies, Pishkin and Shurley (1966a; 1966b) demonstrated a significant inhibiting effect of SI (10–40 hr. of quiet and darkness) on subsequent bar-pressing rate for water reward of a naive, severely water-deprived rat. The presence of another rat, familiar or unfamiliar, during the SI, eliminated the inhibiting effect of SI on subsequent bar-pressing. From these studies, it was expected that SI would inhibit problem solving, as measured by the CI procedure. Therefore, the results reported here were unexpected in showing not inhibition, but enhancement, following the first period (30 min.), then marked inhibition at 60 min., and subsequent improvement again after 120 min., although not significantly different from the 60-min. condition.

The poorer performance with durations beyond 30 min. clearly interacts with complexity of the problem. This is in line with Suedfeld's (1964) result that performance on more complex cognitive tasks was impaired by perceptual isolation, while that on simpler tasks (rote memory and recall of adjective lists) was facilitated.

The preferred explanation for present results is that the markedly reduced external inputs from all modalities, over a 30-min. period, diminish competing and distracting stimuli, permitting *S*s to focus their attention more successfully on the problem. Thus, *S*s perform more efficiently than in the "base-line laboratory" and control HHDE conditions, and this effect is merely accentuated when more complex problems are presented. Furthermore, it is noteworthy that the enhancement of the 30-min. group reflects the effects of actual time spent in HHDE and not the influence of differential stimulation produced in the HHDE, since the control HHDE group (immediate onset of CI) performed at the same level as the standard laboratory group. Thus, it is a safe assumption that the immediate auditory input provided by the CI task is sufficient to detract from both the positive (30 min.) and the negative (60 and 120 min.) effects of HHDE, since the control HHDE group performed at essentially the same level as the standard laboratory *S*s.

After 60 min. a different situation apparently obtains. It is plausible that stimulus flux becomes insufficient to maintain the tonic activation of the ascending reticular activating system. Or the competing stimuli from internal sources (including memory and fantasy) may impair attention. A fairly remote speculation would tie these findings to one of the basic circadian rhythm manifestations in man where a shift from a predominantly conceptual mode of thinking to a predominantly perceptual mode with a periodicity of approximately 60 to 90 min. throughout the night is demonstrated (Dement & Kleitman, 1957). There is some evidence from sleep research that this tendency is also manifested during waking hours, except that it is largely masked by the more potent effects of varied stimulation and motor activity characteristic of waking life (Hartmann, 1968). In the "twilight" type of situation—half-waking and half-sleeping—that is characteristic of SI, this tendency may be more in evidence, with *significantly poorer performance at 60 min.* after onset of SI. Concomitant EEG and other psychophysiological recordings might help to test this hypothesis.

The second major finding of interest in the present experiment relates to the effect on time estimation task. Time of day was controlled for, in that every run began at 1:00 P.M., and *S* was informed that he should keep his entire afternoon free for the experiment. The consistent underestimation of time so regularly found with SI was to assume that time estimation is dependent on a kind of running count kept by the brain of the number of discrete events that happen, and from which a simple additive function would derive a good approximation (Shurley, 1969). Therefore, in SI, the diminished number of discrete events would result in underestimation of elapsed time. These results bring this explanation into question, for the higher number of trials required with increasing problem complexity would be expected to "count" as a higher number of discrete events, and thus diminish, rather than increase, the discrepancy in the two and three irrelevant-dimension problems. On the other hand, if one assumes that *S* in the SI becomes bored from the monotony, then the intellectual challenge of the cognitive task might provide some relief from this condition, with a compensating tendency to reduce overestimation due to monotony and boredom.

REFERENCES

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *Journal of Experimental Psychology*, 1955, **49**, 153–164.

BULGARELLA, R. G., & ARCHER, F. J. Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. *Journal of Experimental Psychology*, 1962, **63**, 254-257.

DEMENT, W., & KLEITMAN, N. Cydic variations in EEG during sleep and their relation to eye movements, body modality, and dreaming. *EEG and Clinical Neurophysiology*, 1957, **9**, 673–690.

HARTMANN, E. The 90-minute sleep-dream cycle. *Archives of General Psychiatry,* 1968, **18**, 280–286.

LILLY, J. C. Mental effects of reduction of ordinary levels of physical stimuli on intact, healthy person. *Symposium No. 2, Group for the Advancement of Psychiatry,* 1956, 13–20.

PISHKIN, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *Journal of Experimental Psychology,* 1960, **59**, 371–378.

PISHKIN, V., & SHURLEY, J. T. Auditory dimensions and irrelevant information in concept identification of males and females. *Perceptual and Motor Skills,* 1965, **20**, 673–683.

PISHKIN, V., & SHURLEY, J. T. Sensory deprivation and operant conditioning of rats. *Psychonomic Science,* 1966, **5**, 283–284. (a)

PISHKIN, V., & SHURLEY, J. T. Social facilitation and sensory deprivation in operant behavior of rats. *Psychonomic Science,* 1966, **6**, 335–336. (b)

ROSSI, A. M., FURHMAN, A., & SOLOMON, P. Arousal levels and thought processes during sensory deprivation. *Journal of Abnormal Psychology,* 1967, **72**, 166–173.

SCOTT, T. H., BEXTON, W. H., HERON, W., & DOANE, B. K. Cognitive effects of perceptual isolation. *Canadian Journal of Psychology,* 1959, **13**, 200–209.

SHURLEY, J. T. Profound experimental sensory isolation. *American Journal of Psychiatry,* 1960, **117**, 539–545.

SHURLEY, J. T. The hydro-hypodynamic environment. *Proceedings Third World Congress of Psychiatry, Montreal* 1961, **3**, 232–237.

SHURLEY, J. T. Stress and adaptation to sensory perceptual isolation research. *Military Medicine,* 1966, **131**, 254–258.

SHURLEY, J. T. Reduced sensory input states. In M. McCally (Ed.), *Hypodynamics and hypogravics.* New York: Academic Press, 1969.

SUEDFELD, P. Toward greater specificity in evaluating cognitive and attitude changes. *American Psychologist,* 1964, **19**, 560. (Abstract)

SUEDFELD, P., GLUCKSBERG, S., & VERNON, J. Sensory deprivation as a drive operation: Effects upon problem solving. *Journal of Experimental Psychology,* 1967, **75**, 166–169.

VERNON, J. *Inside the black room.* New York: Clarkson N. Potter, 1963.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative and Neurological Psychology,* 1908, **18**, 459–482.

ZISKIND, E. An explanation of mental symptoms found in acute sensory deprivation: Researches 1958–1963. *American Journal of Psychiatry,* 1965, **121**, 939–945.

(Received for early publication April 3, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 205–215

INFLUENCE OF LANGUAGE ON SOLVING THREE-TERM SERIES PROBLEMS[1]

HERBERT H. CLARK[2]

Carnegie-Mellon University

Three linguistic processes have been previously proposed to account for the influence of language on deductive reasoning. To test for these processes, 100 *S*s were each asked to solve 64 different forms of the three-term series problem (e.g., "If John isn't as bad as Bill, and Pete isn't as good as Bill, then who is worst?"). The *S*s were allowed only 10 sec. per problem. The kind and relative number of errors *S*s made support the existence of an implicit problem solving strategy containing the three proposed linguistic processes. The strategy consists of (*a*) interpreting the propositions and question of a problem in terms of their deep structure relations, (*b*) retaining the information in this form, and (*c*) later searching memory for information that is congruent in deep structure with the information asked for in the question. For some problems, this strategy leads to longer solution times, hence more errors. For many problems, it also leads to preliminary solutions that show up as specific errors.

People are often led to reason fallaciously merely by the way reasoning problems are worded. Wording is particularly important in the so-called three-term series problems. To cite an example, the problem, "If John is better than Pete, and John is worse than Bill, then who is best?," encourages far more errors in problem solvers than the problem, "If John is better than Pete, and Bill is better than John, then who is best?," even though both problems superficially present the same information (Burt, 1919; DeSoto, London, & Handel, 1965; Hunter, 1957). The immediate aim of the present study was to account for the errors people make in solving such problems. But the wider purpose was to explain the memory processes of storing and retrieving abstract information retained from sentences. In a recently proposed theory for these memory processes (Clark, 1969), it was argued that the way in which a person stores and retrieves the information in a reasoning problem determines how quickly he solves the problem.

The theory proposed by Clark (1969) consists of three principles, briefly summarized here. The first is the principle of the primacy of functional relations. It asserts that the functional relations underlying a sentence, like the logical subject, verb, and object, are more available after comprehension than other less fundamental kinds of information. This principle, derived from linguistic theory (e.g., Chomsky, 1965), was first proposed in slightly different form by Miller (1962) and was supported subsequently by Mehler (1963), McMahon (1963), Gough (1965, 1966), Savin and Perchonock (1965), Slobin (1966), and Clark and Begun (1968).

Comparative sentences, like other constructions, have underlying functional relations. As several linguists (Chomsky, 1965; Doherty & Schwartz, 1967; Huddleston, 1967; Lees, 1961; Smith, 1961) have shown, the sentences "John is worse than Pete" and "Pete isn't as bad as John" both consist linguistically of two "base strings," *John is bad* and *Pete is bad*, that are conjoined in a comparative or equative construction. The first principle, then, asserts that a person "knows" more readily that John and

---

[1] The work reported here was supported in part by United States Public Health Service Research Grant MH-07722 from the National Institute of Mental Health. The author wishes to thank Stuart K. Card and Eve V. Clark for their suggestions on the manuscript and Judy Grumet for her assistance and advice in running the experiment.

[2] Requests for reprints should be sent to Herbert H. Clark, who is now at Stanford University, Department of Psychology, Stanford, California 94305.

Pete are bad (the functional relations expressed in the base strings) than that John is more extreme in badness than Pete.

The second principle, that of lexical marking, specifies the semantic complexity of words like *good* and *bad*. Several generations of linguists (e.g., Bierwisch, 1967; Greenberg, 1966; Lyons, 1963, 1968; Sapir, 1944) have observed that word pairs like good and bad are not symmetrical. Good, the so-called unmarked member of the pair, can be neutralized in some contexts, as in "How good was the movie?," whereas bad, the marked term, cannot. Good, but not bad, can also be neutralized in comparatives: "John is better than Pete" can mean that John and Pete are only being compared evaluatively, although "Pete is worse than John" presumes Pete and John to be bad. The principle of lexical marking is that the neutral senses of unmarked adjectives, like good, are coded in memory in a simpler form than the senses of marked adjectives, like bad. The main consequence is that storage and retrieval should be quicker for comparatives containing unmarked adjectives.

Lastly, the principle of congruence specifies how a search through memory will be made. Information can only be retrieved when it is congruent with the information being sought. Furthermore, this search demands congruence, not of superficial information like words or phrases, but of underlying functional relations. For example, the proposition "John isn't as bad as Pete" implies that John is bad and Pete is bad; the question "Who is best?" requires an X fitting the description X *is good*. Here there is a lack of congruence of underlying functional relations in proposition and question. The question must be implicitly reformulated to read "Who is least bad?" It is only at this point that congruent information can be found and the answer retrieved. Reformulation of a question, of course, should take more time.

A particular strategy for solving problems is implied in the three principles previously described. It goes as follows: The problem solver, in understanding a problem, retains information in such a manner that the underlying base strings are most accessible. (Comprehension time per se is presumably equal for all the comparative sentences used in the present study, and equal for all the negative equative sentences, since the same linguistic constructions are involved in all these comparatives.) It takes him longer in this process to store and retrieve semantically more complex, marked adjectives (like bad) than it does unmarked adjectives (like good). The question directs him then to search for an answer. First, he searches for stored information that is congruent with the question. If he succeeds, he can begin to look at the second-order comparative information and attempt to form an answer; if he fails to find con-

TABLE 1

TYPES OF DETERMINATE THREE-TERM SERIES PROBLEMS

Problem type	Form of problem	Analysis	Problem type	Form of problem	Analysis
I	A better than B B better than C	A is good B is good C is good	I′	A not as bad as B B not as bad as C	A is bad B is bad C is bad
II	C worse than B B worse than A	A is bad B is bad C is bad	II′	C not as good as B B not as good as A	A is good B is good C is good
III	A better than B C worse than B	A is good B is good, bad C is bad	III′	A not as bad as B C not as good as B	A is bad B is bad, good C is good
IV	B worse than A B better than C	A is bad B is bad, good C is good	IV′	B not as good as A B not as bad as C	A is good B is good, bad C is bad

gruent information and a solution, he reformulates the question and continues his search. It is not meant by the term "strategy" that the solver has consciously chosen this means of solving the problem. More likely, the strategy is a normal part of processing language and is used without awareness.

The presence of this strategy can be tested by the many possible forms of the three-term series problem, each of which consists of two propositions followed by a question. Traditionally, this allows 16 possible problems. The four pairs of propositions, numbered I–IV in Table 1, can occur in either order, and the question can be either "Who is best?" or "Who is worst?" These traditional problems, however, can be supplemented by replacing *is better than* by *isn't as bad as* and *is worse than* by *isn't as good as*. The resulting 16 problems are shown on the right of Table 1 as Problem Types I′–IV′. The lettering convention used in Table 1 (and Table 2) is that A is always best, C worst, and B in the middle.

The 32 three-term series problems of Table 1 are related to each other in important ways. First, the problems can be matched for the order in which the three terms appear in the surface structure of the sentences (Chomsky, 1965; Postal, 1964). Problem Type I, e.g., is superficially similar to I′: both types arrange the terms A, B, and C in the same order, and both contain relational terms meaning "strictly greater in goodness than." Types II and II′, III and III′, and IV and IV′ are similar in precisely the same way. But the problems can also be matched for deep structure, roughly the underlying functional relations expressed in the propositions. Problem Type I is built on the functional relations *A is good* and *B is good,* as shown in the Analysis column. In this respect, I is like II′, which is built on the same base strings. Furthermore, Types II and I′, III and IV′, and IV and III′ are similar in deep structure. The most important consequence of the proposed strategy is that deep structure takes precedence over order in surface structure in three-term series problems. This point is implied in all the predictions that follow.

Reasoning by the proposed strategy, a problem solver allowed only 10 sec. per problem should make more errors on some problems than on others. Consider first the problems with two *good* or two *bad* propositions: I, II, I′, and II′. Since the storage and retrieval of information from propositions with *good* should be quicker than from ones with *bad,* the problem solver should make fewer errors overall on I and II′, both problems containing *good,* than on II and I′, both problems containing *bad*. The search stage of the strategy also has its effects on these problems. There is congruence when the two propositions in Problems I and II′ are followed by "Who is best?" and when those in II and I′ are followed by "Who is worst?" The problem solver should make few errors on these combinations, but many errors on the remaining incongruent combinations.

The strategy has important consequences for the problems with heterogeneous propositions (III, IV, III′, IV′) as well. Consider Problem Type III. With the question "Who is best?," the problem solver will search for congruent information and come up immediately with the correct answer A; with "Who is worst?," he will likewise find congruent information in the correct answer C. But consider Problem Type IV. With the question "Who is best?," the search for a congruent solution will produce B (of the second proposition), which, on testing, is clearly wrong. Only after "Who is best?" is implicitly reformulated as "Who is least bad?" is the correct solution (A) directly forthcoming. The same difficulty will also occur with the question "Who is worst?" Type III problems, therefore, should encourage few errors since the strategy leads directly to the correct answer, but Type IV problems should encourage many errors. Type IV′ problems are similar to Type III problems in deep structure (note their identity in the Analysis column), and Type III′ problems are similar to Type IV problems. So, for the same reasons, there should be fewer errors on IV′ than III′.

One kind of problem that has not been studied previously is the so-called indeterminate problem. It is called indeterminate because it contains an incomplete rank order of the three people being compared, as in "John is better than Pete, and Dick is better than Pete." Note that when the question for this pair of propositions is "Who is best?," the correct answer is "can't tell," although when it is "Who is worst?," the correct answer is "Pete." Altogether, there are 16 possible indeterminate problems containing *better* or *worse* or both and 16 others containing *isn't as bad as* or *isn't as good as* or both.

The reasoning strategy proposed here has particularly interesting effects on indeterminate problems. In the problem, "If John is better than Pete, and Pete is worse than Dick, then who is best?," a solver will sometimes make an error, choosing "John" or "Dick" rather than the solution "can't tell." But which term will he choose? The strategy describes the solver as searching first for information congruent with the question. His efforts to satisfy the conditions of the question will lead him to answer "John," but not "Dick": "John" satisfies the congruence conditions, whereas "Dick" does not. It is only in indeterminate problems like this one that the strategy produces errors in such a direct way. Here a congruent and an incongruent solution, both incorrect, can be presented in the propositions in a symmetrical form.

METHOD

Three problems were constructed for each of the 64 possible problem types—the 32 determinate and 32 indeterminate types. In addition, 16 practice problems, 8 determinate and 8 indeterminate, were constructed to have one comparative and one negative equative proposition each. Common four-letter English men's names were used as terms in the problems, with the restriction that no two names could occur together in more than 1 problem. Each problem was printed on an IBM card in the following form:

If Pete isn't as good as Jack,
And Jack isn't as bad as Dick,
Then who is best?
Dick Pete Jack Can't tell

Each *S* solved the 16 practice problems and then 64 other problems, 1 from each of the 64 problem types. Both the practice and experimental problems were in a random and different order for each *S*. There were three groups of *S*s, so that the order of the three names following the question could be counterbalanced both across problems within *S*s and across groups of *S*s within problems; "can't tell" always occurred last among the possible solutions.

Each *S* was given a deck of problems face down. On a signal every 10 sec., he turned a new problem face up, read the problem to himself, circled the answer he thought was correct (or, if he had no answer at the signal, circled nothing), and went on to the next problem. There was a short break after every sixteenth problem. The *S*s, 100 introductory psychology students fulfilling a course requirement, had explained to them beforehand what constituted a "can't tell" answer. They were run in groups of 5–20.

RESULTS

The solutions to both determinate and indeterminate problems contained errors supporting the existence of the reasoning strategy proposed in the present paper. The two kinds of problems are discussed separately. In all that follows, the percentage of errors are compared statistically with Seeger and Gabrielsson's (1968) extension of Cochran's Q test; this extension allows comparisons involving repeated observations on the same *S*.

Determinate problems.—The error data in Table 2 first show that *S*s retained the underlying deep structure relations from the propositions of each problem. Among the homogeneous problems, Type I problems produced fewer errors than Type II problems; this closely replicates the results of DeSoto et al. (1965). Now if the order of the terms in surface structure had been responsible for this difference, then Type I′ problems should have produced fewer errors than Type II′ problems, but if deep structure had been responsible, II′ should have produced fewer errors than I′. Clearly, deep structure takes precedence over surface structure, for I and II′ produced fewer errors, respectively, than II and I′, $\chi^2\ (1) = 24.37$, $p < .001$. The same result holds for Heterogeneous Problems III, IV, III′, and IV′. Type III problems were easier than Type IV problems, again replicating DeSoto et al.) in parallel with this in deep but not surface structure, Type IV′ problems were

TABLE 2

PERCENTAGE OF ERRORS IN SOLVING DETERMINATE THREE-TERM SERIES PROBLEMS

Problem type	Form of problem	Form of question		Average	Overall average
		Best?	Worst?		
I	(a) A better than B; B better than C	15	29	22	17
	(b) B better than C; A better than B	7	18	12	
II	(a) C worse than B; B worse than A	43	48	46	33
	(b) B worse than A; C worse than B	21	21	21	
III	(a) A better than B; C worse than B	8	10	9	10
	(b) C worse than B; A better than B	10	12	11	
IV	(a) B worse than A; B better than C	30	25	28	32
	(b) B better than C; B worse than A	32	43	38	
I′	(a) A not as bad as B; B not as bad as C	28	21	24	31
	(b) B not as bad as C; A not as bad as B	39	36	38	
II′	(a) C not as good as B; B not as good as A	14	32	23	26
	(b) B not as good as A; C not as good as B	26	31	28	
III′	(a) A not as bad as B; C not as good as B	34	45	40	40
	(b) C not as good as B; A not as bad as B	34	45	40	
IV′	(a) B not as good as A; B not as bad as C	26	30	28	28
	(b) B not as bad as C; B not as good as A	23	35	29	

easier than Type III′ problems, χ^2 (1) = 57.30, $p < .001$.

Some of the same results also confirm that unmarked adjectives (like good) are stored and retrieved more quickly than marked adjectives (like bad). Note that the retention of deep structure relations would have been confirmed even if Problem Types I and II′ had been solved with more, rather than fewer, errors than Types II and I′, respectively; the important thing for the confirmation was that I and II′ had the same deep structure relations, and these relations were different from those of II and I′. But since I and II′, containing *good* in their deep structure, *were* solved with fewer errors than II and I′, containing *bad* in their deep structure, this also constitutes evidence for lexical marking; *good* was apparently more quickly comprehended and retrieved than *bad*.

And finally, the data in Table 2 support the proposed search stage of the reasoning strategy. Problem Types III and IV′ are internally congruent: the question always requests information congruent with that specified in the propositions. By the same criterion, Problem Types IV and III′ are internally incongruent. The congruent problems III and IV′ produced significantly fewer errors than the incongruent problems IV and III′. The principle of congruence also finds confirmation in the homogeneous problems. The question "Who is best?" is congruent with the propositions in I and II′, but incongruent with the propositions in II and I′. "Who is best?," rather than "Who is worst?," was answered with fewer errors in I and II′, but with more errors in II and I′, overall χ^2 (1) = 9.04, $p < .01$. In detail, there was one exception to the last result, with "Who is best?" eliciting slightly fewer errors than "Who is worst?" on Problem Type II, a homogeneous *bad* problem. But this is only an apparent exception since "Who is best?," which contains the unmarked adjective *good*, produced fewer incorrect solutions, overall, than "Who is worst?" (24% to 30%). In line with predictions, the difference between *best* and *worst* questions was not as large in Type II problems (2.5%) as it was overall (6%).

It should be noted here that in a previous study (Clark, 1969), *S*s were timed from the beginning of reading to their spoken solutions for each of the 32 determinate problems. The significant differences in errors of the present study are paralleled by significant differences in solution times in the previous study. The problems that produced more errors here had longer solution times previously.

The specific errors *S*s made on these problems further support the existence of the reasoning strategy. The errors can be of three kinds: (*a*) failures to respond, (*b*) "can't tell" answers, or (*c*) incorrect men's names. If *S*s reason as proposed, they should take longer to solve some problems than others. This in turn means that in 10 sec., fewer *S*s will come to any answer whatsoever on the longer problems. And *S*s with no answers will most likely make *a* errors, but also *b* or *c* errors. Indeed, the percentages of *a* errors increased on longer problems. There were 6% such errors on I, compared with 10% on longer II; 10% on II′, compared with 12% on longer I′; 8% on the homogeneous problems (I, II, I′, II′) with congruent questions, compared with 10% on those with incongruent questions; 4% on internally congruent III, compared with 12% on incongruent IV; and 11% on internally congruent IV′, compared with 18% on incongruent III′. The addition of *b* responses merely increases each of these differences, respectively, to 11% to 18%, 16% to 20%, 14% to 18%, 6% to 22%, and 22% to 26%. The further addition of the *c* errors, of course, has been discussed previously.

The *c* errors, incorrect men's names, can indicate the preliminary solutions *S*s tend to choose on some problems or can indicate mere guessing. In determinate problems, the proposed strategy allows only very weak predictions as to what the preliminary solutions will be. In the homogeneous problems (I, II, I′, II′), both of the incorrect men's names should occur, perhaps with the middle B term predominating, because it answers the question when the wrong proposition is considered alone. The results showed here that B occurred more often than the other incorrect term in 12 of the 16 problems, 6% vs. 5% of the time. On the heterogeneous problems (III, IV, III′, IV′), however, B should occur slightly more often than the other term because in III and IV′ B is the only other congruent term and in IV and III′ it is the only congruent term satisfying the question when the wrong proposition is considered alone. The results showed that B was given more often than the other incorrect term on 12 of the 16 problems, 6% vs. 3% of the time.

Indeterminate problems.—Errors in indeterminate three-term series problems also give evidence that *S*s reason by the proposed strategy. Table 3 lists the percentages of errors for each type of problem making use of a new lettering convention in which J is the repeated term and G and H are arbitrarily assigned to the other two terms.

The data here again show that *S*s retain underlying functional relations from the problems they have read: deep structure, rather than order of terms in the surface structure, predicts the differences in these problems. Problem Types V and VI′, which have the same deep structure, were solved with fewer errors than Types VI and V′, respectively, $\chi^2\ (1) = 4.85,\ p < .05$. The similarity of surface structure between V and V′ and between VI and VI′ had no perceptible effect on their solutions. The heterogeneous problems VII, VIII, VII′, and VIII′ cannot be compared in the same way since whether the question was answerable or not is confounded with surface structure; answerability will be shown subsequently to be very important in the production of errors.

The effect of lexical marking—making *good* more accessible than *bad*—was again found when Problem Types V and VI′, with underlying *good*, were solved significantly more easily than Problem Types VI and V′, with underlying *bad*.

The strategy of searching for congruent information derives its best and most direct support from the problems with heterogeneous propositions (VII, VIII, VII′, VIII′ in Table 3). Consider the problem, "If D is better than E, and E is worse than F, then who is best?," whose correct solution is "can't tell." (In a temporary lettering convention for this kind of problem, the D term is congruent with the question, the F

TABLE 3

Percentage of Errors in Solving Indeterminate Three-Term Series Problems

Problem type	Form of problem	Form of question		Average	Overall average
		Best?	Worst?		
V	(*a*) G better than J; H better than J	4[a]	15	10	9
	(*b*) J better than G; J better than H	13	4[a]	8	
VI	(*a*) G worse than J; H worse than J	20	14[a]	17	14
	(*b*) J worse than G; J worse than H	7[a]	14	10	
VII	(*a*) G better than J; J worse than H	49[a]	36	42	42
	(*b*) J worse than H; G better than J	45[a]	37	41	
VIII	(*a*) J better than G; H worse than J	22	44[a]	33	35
	(*b*) H worse than J; J better than G	23	52[a]	38	
V′	(*a*) G not as bad as J; H not as bad as J	15[a]	28	22	24
	(*b*) J not as bad as G; J not as bad as H	28	25[a]	26	
VI′	(*a*) G not as good as J; H not as good as J	18	14[a]	16	19
	(*b*) J not as good as G; J not as good as H	14[a]	29	22	
VII′	(*a*) G not as bad as J; J not as good as H	66[a]	42	54	56
	(*b*) J not as good as H; G not as bad as J	66[a]	49	58	
VIII′	(*a*) J not as bad as G; H not as good as J	35	61[a]	48	45
	(*b*) H not as good as J; J not as bad as G	28	56[a]	42	

[a] The correct solution is "can't tell" to these problems and "J" to the rest.

term is incongruent, and the E term is the repeated term.) When *S*s make mistakes on this problem, they should answer the "half correct" D and F more often than E. Moreover, the proposed reasoning strategy predicts that in searching for information congruent with the question, *S*s will choose the D term as their preliminary solution. In agreement with this prediction, the results showed that all eight heterogeneous problems with "can't tell" answers produced considerably more congruent D errors than incongruent F errors ($p < .01$, by a sign test). The D errors occurred, on the average, 26% of the time, E errors 3%, and F errors 9%. The ratio of D to F was almost 3:1; on the four appropriate practice problems, the ratio was 4:1.

For each pair of propositions in Table 3, one question is answerable with a name and the other is unanswerable, i.e., the solution is "can't tell." The difficulties of answerable and unanswerable problems seemed somewhat paradoxical. When the propositions were homogeneous, each of the eight answerable problems produced more errors than its correlative unanswerable problem ($p < .01$, by a sign test). On the other hand, when the propositions were heterogeneous, each of the eight answerable problems produced *fewer* errors than its correlative unanswerable problem ($p < .01$, by a sign test). Both effects, it is argued, constitute evidence for the search for congruence as a stage in the reasoning strategy.

First, in the homogeneous problems (V, VI, V′, VI′), it is easy, in comprehending the two propositions one after the other, to compare the information in them, for the second proposition is exactly congruent with the first except in the first or last term (cf. V or VI in Table 3). By comparing these two almost identical statements, then *S* immediately knows that the three-term series is indeterminate and that the solution might well be "can't tell." If time is running short, therefore, he will give his preliminary solution of "can't tell"; otherwise he will work towards the correct answer. This tendency to answer "can't tell," however, leads to correct solutions on unanswerable problems, whose correct solution *is* "can't tell," but incorrect solutions on answerable problems. The results confirm this prediction. The *S*s made errors on 19% of the answerable problems but on only 12% of the unanswerable ones. Yet they failed to respond equally often (5% of the time) on both. Most of the remaining errors on the answerable problems were the predicted "can't tell" answers, which were produced on 13% of the problems, thus accounting for the greater number of errors in answerable problems than in unanswerable problems.

In the heterogeneous problems (VII, VIII, VII′, VIII′), on the other hand, it is very difficult to compare the information in the two propositions, for the second proposition is completely incongruent with the first. Consider the pair of propositions, "Jack isn't as bad as Pete, and Pete isn't as good as Dick." When the question is "Who is worst?," *S* will search and find a preliminary congruent solution in the "Pete" of the first proposition. That solution also happens to be correct. When the question is "Who is best?," he will search and find a preliminary congruent solution in "Dick." But this solution happens to be incorrect. The search for congruent information, then, leads to correct preliminary solutions on answerable problems but to incorrect ones on unanswerable problems. This prediction from the proposed reasoning strategy is confirmed by the errors. The *S*s made errors on 34% of the answerable problems but on 55% of the unanswerable ones. Nevertheless, they failed to respond equally often (17% of the time) on both. The remaining 17% of the errors for the answerable problems were more or less evenly distributed among the other possible errors. But the remaining 38% of the errors for the unanswerable problems were not: *S*s produced the congruent, but incorrect, answer on fully 26% of the unanswerable problems, as noted previously, thus accounting for the greater number of errors in the unanswerable problems than in the answerable ones.

Order of two propositions.—The present *S*s also appeared to make use of a memory-saving strategy in which the information of each proposition was compressed into a less cumbersome form. Instead of storing all the information of a first proposition, "John is better than Bill," they stored only an abbreviated version which might be glossed as "John is the better one." If this *John* then appeared in the second proposition, e.g., "Pete is better than John," alignment of the three terms was relatively easy, for Pete is better than John who was already the better one. But if this *John* did not appear in the second proposition, then alignment was not immediate. The *S*s had to backtrack for the full version of the first proposition or revert to some other time-consuming strategy.

The strategy of compressing information, in general, makes one ordering of two propositions easier than the other (see Clark, 1969). In Problem Types I and II, the *b* order should be easier than the *a* order since the *b* order allows the compressed version of the first proposition to be used in aligning the terms of the second. This prediction was borne out in the data, χ^2 (1) = 31.02, $p < .001$. But the compressed version of the proposition, "Bill isn't as good as John," is also implicitly "John is the better one." In this case, it is the second term, and not the first, that serves as the term in the compressed version. So in Problem Types I′ and II′, it is the *a* order that should be the easier order. This prediction, too, was borne out in the data, χ^2 (1) = 7.58, $p < .01$. The two orderings of the propositions of III, IV, III′, and IV′, on the other hand, do not differ with the application of the compressed strategy, and, except for IV, the data show only slight differences in errors between the two orders. Note here that the compressed version of a proposition is not directly related to its surface structure: it is neither the subject nor predicate term which serves as the focus of the compressed information, but the term more extreme on the underlying goodness or badness scale.

The compression strategy also affected the error rates in the indeterminate problems. In each of the eight Roman-numeraled pairs of problems listed in Table 3, one problem was favorable to the application of the compression strategy and the other was not. In all eight pairs, the favorable problem produced fewer errors than the unfavorable problem, ($p < .01$, by a sign test), although the differences averaged only 5%.

Discussion

There is good evidence in the present results for all three stages of the proposed strategy for deductive reasoning. First, in comprehending the propositions of a problem, *S*s retained underlying functional relations like *John is bad*. The evidence for this was that the relative number of errors for two reasoning problems was predictable in each case from the deep structure, not the superficial characteristics, of the problem's propositions. Furthermore, deep structure congruence, rather than incongruence, between a question and the information it asked for resulted in fewer errors in a fixed interval of time. Both kinds of results are evidence for the reality of underlying functional relations in reasoning.

Once a person has stored the abstract elements underlying a sentence, he can later retrieve that information. A priori, such a search must have principles to guide it, for it must decide when it has found the object it is seeking. One of the few reasonable assumptions is that the search terminates when it finds information congruent to the information sought. This congruence must be at an abstract level. In three-term series problems, the underlying functional relations of the question have to be congruent with those of the propositions. In this respect, "Who is best?" requires information congruent with the information in "Jack isn't as good as Pete." Several strong tests of this principle in the present experiment were all confirmatory.

One final aspect of the reasoning strategy is that some kinds of information are more easily stored and retrieved than others. The abstract semantic features of unmarked adjectives, like good, are less complex than those of marked adjectives, like bad, and so it was expected that retrieval should reflect this complexity. Four tests of this principle for good and bad all showed that retrieval time did reflect semantic complexity. This evidence alone would not be enough to support the principle of lexical marking, for good–bad is only one such unmarked–marked pair. But previous experiments (e.g., Clark, 1969; DeSoto et al., 1965; Handel, DeSoto, & London, 1968; Hunter, 1957; Huttenlocher, 1968) have required *S*s to solve problems with homogeneous propositions (problems of Types I and II) containing better–worse, more–less, faster–slower, farther–nearer, happier–sadder, warmer–colder, taller–shorter, and deeper–shallower. The problems containing an unmarked comparative, the first member of each pair, were solved more quickly or with fewer errors in a fixed interval of time than the problems containing its marked counterpart. Other evidence for lexical marking is discussed in Clark (1969).

The strategy proposed here, however, predicts more than just the *number* of errors *S*s should make on three-term series problems: in many instances it also predicts *which* errors they should make. Since *S* searches his memory for underlying base strings congruent with those of the question, he should sometimes come to preliminary solutions that fulfill the congruence conditions, but which are nevertheless incorrect. Specific errors confirming these predictions were found, especially in the indeterminate problems. These errors demonstrate that in some problems *S*s try at least one incorrect solution before they come to the correct one and that the ones they try are determined by their search strategy.

The present results are in disagreement with several previous theories of the solution of three-term series problems (DeSoto et al., 1965; Hunter, 1957; Huttenlocher, 1968). According to previous results (Clark, 1969), each of these theories incorrectly predicts the relative solution times of certain three-term series problems, especially those with "isn't as good as" and "isn't as bad as" propositions. The present error results, because they fully parallel these solution times, constitute further disconfirmation of these past theories (see Clark, 1969, for details).

Another older theory (Chapman & Chapman, 1959; Sells, 1936; Woodworth & Sells, 1935) asserts that there is an "atmosphere effect" that accounts for certain difficulties in reasoning. This effect occurs, e.g., in traditional syllogisms that contain two negative premises (or propositions). The negatives are said to induce a "negative atmosphere" in this problem, causing the solver to draw a negative, hence often invalid, conclusion. The present three-term series problems have difficulties that also look like the result of a subtle atmosphere effect. But this effect can only *describe* what happens; by itself, it explains nothing. The present proposal instead attempts to account for these difficulties by general linguistic processes. Neither the "atmosphere effect" nor a similar descriptive statement can be said to explain the difficulties in that sense. On the contrary, the atmosphere effect itself may eventually be explained by linguistic processes similar to the ones proposed here to account for solving three-term series problems.

Finally, the present study appears to put important limitations on general theories of linguistic memory. The emphasis here has been on the fact that what a person retains from a sentence is highly abstract, even though he uses the information immediately. Memory for linguistic material over longer periods of time is also highly abstract in form, as has been shown by Brown and McNeill (1966), Fillenbaum (1966, 1968), Sachs (1967), Clark and Clark (1968), Henley, Noyes, and Deese (1968), Clark and Stafford (1969), and Clark and Card (in press), as well as by those mentioned previously, as confirming the importance of deep structure. It is difficult to see how long-term memory could be otherwise, given the present results. If what a person retains for immediate use is so abstract, he must certainly retain the same kind of information over longer periods of time. The more concrete attributes of a sentence, its words, phrases, or surface structure, appear to be done away with soon after its having been heard, unless there is a special reason to keep them. What remains is an abstract interpretation for use either immediately or later.

REFERENCES

BIERWISCH, M. Some semantic universals of German adjectivals. *Foundations of Language,* 1967, **3**, 1–36.

BROWN, R., & MCNEILL, D. The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 325–337.

BURT, C. The development of reasoning in school children. *Journal of Experimental Pedogogy,* 1919, **5**, 68–77 and 121–127.

CHAPMAN, L. J., & CHAPMAN, J. P. Atmosphere effect re-examined. *Journal of Experimental Psychology,* 1959, **58**, 220–226.

CHOMSKY, N. *Aspects of the theory of syntax.* Cambridge: M.I.T. Press, 1965.

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review,* 1969, **76**, 387–404.

CLARK, H. H., & BEGUN, J. S. The use of syntax in understanding sentences. *British Journal of Psychology,* 1968, **59**, 219–229.

CLARK, H. H., & CARD, S. K. The role of semantics in remembering comparative sentences. *Journal of Experimental Psychology,* 1969, in press.

CLARK, H. H., & CLARK, E. V. Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology,* 1968, 20, 129–138.

CLARK, H. H., & STAFFORD, R. A. Memory for semantic features in the verb. *Journal of Experimental Psychology,* 1969, **80**, 326–334.

DESOTO, C., LONDON, M., & HANDEL, S. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology,* 1965, **2**, 513–521.

DOHERTY, P., & SCHWARTZ, A. The syntax of the compared adjective in English. *Language,* 1967, **43**, 903–936.

FILLENBAUM, S. Memory for gist: Some relevant variables. *Language and Speech,* 1966, **9**, 217–227.

FILLENBAUM, S. Sentence similarity determined by a semantic relation: The learning of "converses." *Proceedings of the 76th Annual Convention of the American Psychological Association,* 1968, 3, 9–10.

GOUGH, P. B. Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 107–111.

GOUGH, P. B. The verification of sentences: The effects of delay on evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 492–496.

GREENBERG, J. H. *Language universals.* The Hague: Mouton, 1966.

HANDEL, S., DESOTO, C., & LONDON, M. Reasoning and spatial representation. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 351–357.

HENLEY, N., NOYES, H., & DEESE, J. Semantic structure in short-term memory. *Journal of Experimental Psychology,* 1968, **77**, 587–592.

HUDDLESTON, R. D. More on the English comparative. *Journal of Linguistics,* 1967, **3**, 91–102.

HUNTER, I. M. L. The solving of three-term series problems. *British Journal of Psychology,* 1957, **48**, 286–298.

HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. *Psychological Review,* 1968, **75**, 550–560.

LEES, R. B. Grammatical analysis of the English comparative construction. *Word,* 1961, **17**, 171–185.

LYONS, J. *Structural semantics.* Oxford: Blackwell, 1963.

LYONS, J. *Introduction to theoretical linguistics.* Cambridge: Cambridge University Press, 1968.

MCMAHON, L. E. Grammatical analysis as part of understanding a sentence. Unpublished doctoral dissertation, Harvard University, 1963.

MEHLER, J. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 346–351.

MILLER, G. A. Some psychological studies of grammar. *American Psychologist,* 1962, **17**, 748–762.

POSTAL, P. Underlying and superficial linguistic structure. *Harvard Educational Review,* 1964, **34**, 246–266.

SACHS, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics,* 1967, **2**, 437–442.

SAPIR, E. Grading: A study in semantics. *Philosophy of Science,* 1944, **11**, 93–116.

SAVIN, H. B., & PERCHONOCK, E. Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 348–353.

SEEGER, P., & GABRIELSSON, A. Applicability of the Cochran Q test and F test for statistical analysis of dichotomous data for dependent samples. *Psychological Bulletin,* 1968, **69**, 269–277.

SELLS, S. B. The atmosphere effect: An experimental study of reasoning. *Archives of Psychology,* 1936, **16**, 1–83.

SLOBIN, D. I. Grammatical transformations in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 219–227.

SMITH, C. S. A class of complex modifiers. *Language,* 1961, **37**, 342–365.

WOODWORTH, R. S., & SELLS, S. B. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology,* 1935, **18**, 451–460.

(Received January 10, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 216–223

CRITERION CHANGE AND RESPONSE COMPETITION IN UNLEARNING[1]

WILLIAM P. BANKS[2]

University of California, Berkeley

The extinction hypothesis of retroactive inhibition (RI) was evaluated in two experiments using the A-B, A-C paired-associate paradigm. In Exp. I, modified-modified free recall (MMFR) was augmented with confidence ratings so that original learning (OL) strength (d′) could be assessed at two levels of interference (2 or 20 trials of interpolated learning, IL). The OL d′ was the same for both interference conditions, but OL response criterion (beta) became more restrictive after 20 trials of IL. Three levels of monetary payoff were administered at recall and had no effect on emission of correct responses. In Exp. II, *S*s were forced to respond with at least a guess about every OL and IL response in MMFR and exhibited less RI when so forced. It was concluded, contrary to the extinction hypothesis, that interference changes OL criterion, not strength, but that this criterion change is itself a consequence of generalized response competition.

According to the two-factor theory of interference, retroactive inhibition (RI) in the A-B, A-C paired-associate (PA) paradigm is partly the result of competition of responses at recall and partly the result of unlearning of responses from original learning (OL) during acquisition of the responses of interpolated learning (IL). The modified-modified free recall (MMFR) technique introduced by Barnes and Underwood (1959) is designed to minimize the effects of response competition at recall, and consequently, any forgetting of OL responses in MMFR is ascribed to unlearning. Since its introduction, the MMFR technique has been used in a large number of studies and has given results consistent with an unlearning factor (see Keppel, 1968).

The present study is concerned with the mechanism behind unlearning. Unlearning was originally conceived of as extinction of individual OL responses during IL (Melton & Irwin, 1940), but the strict Pavlovian analogy has not found complete support. In 1956, Newton and Wickens suggested that RI may result from a form of competition operating at a more general level than the individual PA. They felt that "*S*s may acquire a tendency to give a certain class of responses [those from a given list] in addition to giving specific responses to specific stimuli [p. 153]." Thus, interpolation could reduce the availability of OL as a whole by disrupting the tendency to emit responses from that list. Recent experiments by Postman and his associates (Postman, 1967; Postman, Stark, & Fraser, 1968) have supported generalized competition as an important source of RI.

The mechanism of RI investigated in the present experiments is one that is more in accord with the hypothesis of generalized competition than is extinction. It is assumed that *S* attempting recall is performing a search for what it was he learned and that he must choose his response from a number of alternatives. The *S*'s willingness to commit himself in making these choices is, consequently, a factor operating in addition to "strength of trace" to determine amount recalled. Measured forgetting can result either from a loss of trace strength or from an unwillingness to guess. It is hypothesized that IL does not reduce the strength of OL as the extinction analogy implies, but that through generalized competition, it increases the difficulty of choice in OL recall and inhibits guessing, producing a decrement in recall. Both loss of strength and restricted guessing may, of course, be operating, but the present concern is with establishing the guessing factor.

---

[1] This research was conducted while the author was supported by a National Institute of Mental Health predoctoral training grant administered through the Department of Psychology of Johns Hopkins University. The writing was done at the Institute of Human Learning, University of California at Berkeley, under a United States Public Health Service postdoctoral traineeship (5T01-GM01207-05). The Institute of Human Learning is supported by grants from the National Science Foundation and the National Institutes of Health. The author is grateful to James Deese for his excellent advice on all phases of this research and to Norman J. Slamecka for his critical reading of much of the manuscript.

[2] Requests for reprints should be sent to William P. Banks, who is now at Pamona College, Department of Psychology, Claremont, California 91711.

The test of this hypothesis is based on signal detectability theory (SDT). SDT analyzes reported detection of a signal with two parameters: d', the strength of the signal, and beta, the relative strictness or laxness of the "guessing" or decision criterion (Swets, Tanner, & Birdsall, 1961). Using techniques developed by Egan and others (Egan, 1958; Pollack & Decker, 1958), it is possible to analyze memory performance in terms of these parameters. In analogy to psychophysical detection, the memory trace is conceived of as the signal and the strength of the trace (d'), and the decision criterion (beta) for accepting or rejecting it are derived from recall or recognition performance. The technique employed here is based on confidence ratings of responses emitted in recall and is described in the previous references, as well as in a number of learning studies that have used it (e.g., Murdock, 1966). Given the number of correct and incorrect responses in each category of confidence, Type II memory operating characteristic (MOC) plots can be obtained and d' and beta can be determined.

The two experiments reported here measure retention with a MMFR technique that differs in two ways from Barnes and Underwood's (1959). First, Barnes and Underwood's instruction that *S* should try to pair his recalled responses with the stimuli they were associated with in learning is not used, although the recall sheet is set up so that *S*s can attempt correct pairings if they wish, and *S*s were asked only to try to designate correctly the list membership of their responses. This alteration in the Barnes and Underwood instructions was intended to reduce still further the possibility of response competition in MMFR and thus to insure further that whatever RI is found is a result of unlearning and not response competition. It is possible that with the Barnes and Underwood instructions an *S* might withhold a response because he could not decide with which stimulus it had been paired. If OL unavailability and not associative confusion is the essential characteristic of unlearning (see Postman, 1967), then to measure unlearning we must collect all recalled responses, whether or not *S* is sure he knows their proper associates. The second alteration in the Barnes and Underwood procedure is the augmentation of MMFR with confidence ratings so that d' and beta for OL can be estimated after different degrees of IL. The *S*s were instructed to use the confidence categories to rate their certainty that their responses at recall did, indeed, come from the list, OL or IL, to which they assigned them. Under this instruction, the resulting MOC functions should measure availability since list identification is quite accurately made for all recalled responses.

It is predicted that in these experiments beta for OL will become more restrictive as interference increases and that d' will be relatively unaffected. Experiment I includes payoff manipulations to vary beta experimentally. If unlearning is the result of criterial changes, payoff variations should tend to mimic or compensate for the effects of interference. In Exp. II, *S*s are forced to respond to every stimulus in MMFR in a further attempt to vary criterion experimentally.

EXPERIMENT I

Method

Design.—All *S*s learned a PA list to a criterion of one perfect recitation. They then entered one of six independent groups, defined by the factorial combination of two levels of interference and three payoff schedules. Interference consisted of either 2 trials or 20 trials of IL on an A-C list. Payoff was based on the following MMFR scoring: Correct responses had to have list membership, OL or IL, correctly designated, regardless of pairing with OL or IL stimuli. Incorrect responses were all extralist intrusions as well as OL responses designated as IL, or vice versa. The three payoff schedules were: (*a*) 0/0, control condition with no payment or charge; (*b*) 20/0, lenient condition where all correct responses were rewarded with 20¢ and no charge for incorrect ones; and (*c*)

20/20, strict condition where correct responses were rewarded with 20¢ and incorrect ones cost 20¢.

An additional variable, within-*S*s, was the administration of two MMFR tests. A first MMFR (MMFR$_1$) was administered to all *S*s immediately after IL. Confidence rating was required in MMFR$_1$, but no mention was made of payoff. After MMFR$_1$, the sheet was taken away and a second one (MMFR$_2$) was presented. With MMFR$_2$ came payoff instructions; confidence rating was the same as for MMFR$_1$.

Subjects.—There were 48 male Hopkins undergraduates, 8 in each of the six conditions, fulfilling a requirement of an introductory psychology course. Some had previously served in one verbal learning experiment, but all were naive to paired associates. Two additional *S*s were rejected for failure to learn the first list in 40 trials.

Lists.—Both OL and IL had 8 paired associates, CVC stimuli and bisyllabic adjective responses, given at a 2:2 rate with a 4-sec. intertrial interval. Stimuli and responses were selected and paired according to the rules of Barnes and Underwood (1959). Every list was presented in four counterbalanced orders. After 24 *S*s had been tested, new pairings of stimuli and responses were given to the second 24 *S*s.

Procedure.—Learning of OL was by anticipation at a 2:2 rate to criterion. There followed a 1-min. rest period during which IL instructions were given. After IL, there was another 1-min. rest period which was largely taken up with MMFR$_1$ instructions, and MMFR$_1$ was administered immediately after the rest. There was no filler task to equate the OL retention interval for the 2- and 20-trial groups. Recall was written on a sheet on which all stimuli were in a column on the left-hand margin. After each stimulus, two lines were provided for the OL and IL responses. The lines formed two columns, and at the top of the left- and right-hand columns, "First List" and "Second List," respectively, were printed. The *S*s were not required to pair stimuli and responses appropriately, but were instructed only to write each response in the correct column. Below the space for recall of each word was a line for recording confidence judgments. It had seven intervals and was labeled "Most Confident" at one end and "Least Confident" at the other. The *S*s were told to use this scale to rate their confidence of the list membership they gave to each response. Checking "Most Confident" indicated that they were certain they had put the word in the correct list, and "Least Confident" meant a pure guess.

The *S*s were given 2 min. (or more if they requested) to complete MMFR$_1$. Guessing and full use of time were encouraged. After MMFR$_1$ the recall sheet was taken away and a fresh sheet was presented for MMFR$_2$. Instructions for MMFR$_2$ varied according to the payoff condition. The *S*s in the 0/0 group were instructed to fill the sheet out exactly as before, recalling as much as they could, whether or not they recalled it before. In the 20/0 group, *S*s were also told to fill the sheet out as before, but were advised that since there was no penalty for incorrect responses and a reward for correct ones, they would probably increase their gains by guessing after emitting all responses they were sure of. The 20/20 group was advised to be very cautious since they would be more likely to lose than gain by guessing. For all payoff conditions the recall sheet for MMFR$_2$ was identical to the first, and the confidence rating procedure, instructions (except for payoff instructions), and time limit were also the same.

Results

First-list learning.—Mean trials to OL criterion was 15.23, and analysis showed a significant difference between the groups, F $(5, 42) = 2.48$, $p < .05$. A concentration of fast learners in the two nonpayoff groups (0/0 at 2 and 20 trials) was responsible for the difference. With the two 0/0 conditions removed, the remaining four did not differ in OL trials, F $(3, 28) = .59$. Recall data from MMFR$_1$ were analyzed to determine whether OL differences had any effect on recall. Since MMFR$_1$ was taken before payoff, possible differences due to learning speed were not yet contaminated by payoff effects. The payoff groups did not differ significantly in MMFR$_1$ recall, F $(2, 42) = .18$, and there were no significant interactions involving the payoff variable. Dunett's t was applied to the differences between first MMFR means of the 0/0 groups and the corresponding means of the 20/0 and 20/20 groups combined and no differences were significant. On the basis of these analyses, it was decided that OL differences did not contaminate the results. Since payoff had, in fact, no consistent effect on recall in MMFR$_2$ (see subsequent discussion), the question of a confounding of learning differences with the payoff variable does not arise. The MOC analysis was performed after data had been summarized across payoff groups within interference conditions and yielded d' and beta only for the two interference conditions as a whole, between which there was no difference in OL rate.

Response recall.—Table 1 presents mean MMFR recall of OL and IL for each experimental condition. OL recall was analyzed using trials of IL and payoff group as between-*S* variables and MMFR$_1$ vs. MMFR$_2$

TABLE 1

MEAN RECALL OF OL AND IL IN $MMFR_1$ AND $MMFR_2$ AFTER 2 AND 20 TRIALS OF IL IN THREE PAYOFF CONDITIONS (EXP. I)

Payoff	2 trials				20 trials			
	OL		IL		OL		IL	
	$MMFR_1$	$MMFR_2$	$MMFR_1$	$MMFR_2$	$MMFR_1$	$MMFR_2$	$MMFR_1$	$MMFR_2$
0/0	6.87	6.63	3.13	3.25	2.87	3.63	7.75	7.63
20/0	6.87	6.75	3.00	3.37	3.50	4.00	8.00	7.87
20/20	7.25	7.37	3.13	3.00	3.37	3.87	7.25	7.25
M	7.00	6.92	3.08	3.21	3.25	3.83	7.66	7.58

as a within-*S* variable. The oft-replicated unlearning effect was highly significant, F (1, 42) = 67.00, $p < .001$. Payoff seemed to have no influence whatever on recall. The main effect of payoff as well as the interactions involving this variable all had F ratios less than unity. Separate t tests compared the change in recall from $MMFR_1$ to $MMFR_2$ for the 20/0 and 20/20 groups (which did not differ in OL rate) in order to test the possibility that OL differences had masked payoff effects. Although these two groups represented the extremes of payoff and were equated for learning rate, they did not differ significantly in amount of change in recall of OL or IL from $MMFR_1$ to $MMFR_2$ at either level of interference.

The MMFR variable, intended primarily as a vehicle for payoff treatments, interacted significantly with IL trials, F (1, 42) = 6.85, $p < .025$. Inspection of the data reveals that recovery of OL after heavy interference is responsible for this effect.

Stringent scoring.—With stringent scoring, responses are counted as correct only if they are paired in MMFR with the stimuli that elicited them in learning. Although *S*s were not instructed to make correct pairings, it was possible for them to do so, and it was also possible to score all the protocols stringently. In experimental situations very similar to the present one, both Barnes and Underwood (1959) and McGovern (1964) found little difference between the number of responses recalled (lenient scoring) and the number correct by stringent scoring, but they instructed *S*s to attempt correct pairings. In this experiment, without instructions to pair correctly, stringent OL scores were somewhat lower than lenient OL scores, but this difference changed little as interference increased. For OL after 2 trials of IL the average stringent score was 78% of the lenient score, and after 20 trials it had dropped to 63% of the lenient score. Since the level of lenient recall in the present experiment is equivalent to that found in the conditions of the Barnes and Underwood and the McGovern experiments that were essentially replicated, it can be assumed that the somewhat lower stringent scores found here are a result only of the different instructions. The *S*s could have produced higher stringent scores if they had tried, and the data seem to show that the overall stringent scores merely reflect the fact that some *S*s tried to make correct S-R pairs, and did, while the rest did not try at all. For OL recall, 47% of the *S*s have stringent scores equal to their lenient scores, while 23% have stringent scores as bad as or worse than what would be expected by a random pairing of what they recalled with the stimuli.

Intrusions.—Intrusion rates were quite low in all conditions. The maximum possible number of intrusions was 511, but only 64 were obtained. Of these, only 1 was the result of confusion between OL and IL, and the rest were extralist.

Although payoff had some effect in the expected direction on the number of intrusions, the effect was very small, and differences between *S*s in intrusion rate were too great to allow any meaningful statistical tests. The small effect of payoff on intrusion rates is illustrated by performance in the 20/0 group. With this lenient payoff, gains would be maximized by using the lowest criterion possible, i.e., by withholding no responses whatever, and *S*s were made aware of this. Such a strategy should produce either a correct response or an intrusion at every opportunity to respond, yet there were 53 instances when no response at all was given (256 total opportunities, 178 correct responses, and 25 intrusions).

Confidence rating.—Confidence ratings showed a high degree of discrimination between correct and incorrect responses (also see Murdock, 1966; Pollack & Decker, 1958). Correct responses were distributed in a highly skewed curve, with an overall 95% in the category of greatest confidence. Intrusions showed a less extremely skewed distribution, with 67% falling in the three lowest categories overall, and considerable variation existing from condition to condition.

The only strong effect of payoff was on the confidence level of intrusions emitted. Before payoff, at $MMFR_1$, the mean confidence level of the intrusions on a 1.00–7.00 scale (7.00 being the highest) was 3.13, and each *S* emitted an average of .48 intrusions. In the 0/0 control condition, the mean confidence of the intrusions in $MMFR_2$ was about the same; 3.73 and .69 per *S* were emitted. With lenient 20/0 payoff, the mean confidence level dropped to 1.76 as quite a few more intrusions, 1.56 per *S*, were elicited. The strict 20/20 payoff, on the other hand, seemed to suppress intrusions of about the same; 3.73 and .69 per *S* were emitted, and the mean confidence of these was 5.40. The average confidence level of the correct responses was 6.84 in $MMFR_1$. In $MMFR_2$ it was 6.91 for the 0/0 payoff group, 6.82 for the 20/0 payoff, and 6.89 for the 20/20 payoff.

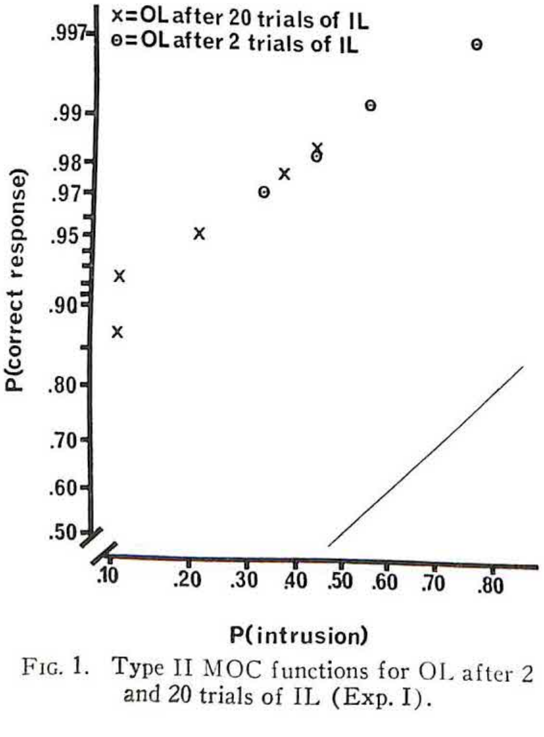


FIG. 1. Type II MOC functions for OL after 2 and 20 trials of IL (Exp. I).

SDT analysis.—For two reasons it was not possible to make meaningful SDT analyses for each condition as was planned. First, the common finding that few intrusions occur in MMFR also obtained here, even when *S*s were highly motivated to make intrusions. In 8 of the 24 cells there were no intrusions at all and consequently no way to determine the noise distribution. Secondly, *S*s were very accurate and extreme in their confidence judgments and placed the few intrusions they emitted in such low categories that, in general, estimates of false-alarm probabilities were based on too few intrusions to be reliable.

Although the more specific tests of the hypothesis could not be made, it was still possible to make a general test by determining the two SDT parameters for OL at the two levels of interference after collapsing over the payoff and MMFR variables. This procedure provided an estimate of d' and beta based on two measures for each *S* ($MMFR_1$ and $MMFR_2$), the second being taken at one of the three different motivational levels. There seems to be no way this procedure could introduce a bias in the measurement of d' for OL in either interference group. At most it increases the range of criteria over which d' is determined. Whatever bias this procedure may exert on measurements of beta affects both OL groups equally, since both are collapsed over the same motivational levels. Accordingly, Fig. 1 presents the MOC functions for OL at the two levels of interference. Each set of points is fitted well by a straight line running nearly parallel to the positive diagonal on probability coordinates. It is clear that sensitivity is very nearly the same for both

TABLE 2

MEAN RECALL AND TOTAL INTRUSIONS OBTAINED IN $MMFR_1$ AND $MMFR_2$ OF EXP. II

Item	2 trials of IL				20 trials of IL			
	OL		IL		OL		IL	
	$MMFR_1$	$MMFR_2$	$MMFR_1$	$MMFR_2$	$MMFR_1$	$MMFR_2$	$MMFR_1$	$MMFR_2$
Recall	7.750	7.625	3.000	3.000	4.000	5.250	7.750	7.875
Intrusions	0	3	4	40	0	22	0	1

levels of interference whatever measure of sensitivity is used since both sets of points fall on the same MOC. At the highest confidence level, d' is 2.36 for OL after 20 trials of interference and 2.32 for 2 trials. The mean d' determined by the five points obtained from the 20-trials group is 2.35, and it is 2.24 averaged over the four points provided by the 2-trials group.

The two functions represent, on the other hand, quite different levels of response criterion. If the two sets of points are assumed to lie on the same MOC, those from the 20-trials group define a much more cautious region than those from the 2-trials group. There is overlap between the two distributions, but the most cautious criterion used after 2 trials of interference is still more lenient than the three strictest criteria used after 20 trials of interference. Mean beta for the three strictest criteria after 20 trials is .83, but only .13 for the three strictest criteria after 2 trials. This is a large difference, although there is no way of testing its significance.

The SDT analysis implies that OL loss after interference is the result of criterion shift and not unlearning. If this is the case, then RI should be eliminated if the criterion is lowered to its preinterference level. Monetary profit was not found to change criterion appreciably in Exp. I, and a more effective technique seemed necessary. Consequently, in Exp. II, a forced responding technique was employed in which *S*s were required to give a response on every blank on the $MMFR_2$ sheet. Forced responding should induce the lowest possible response criterion in *S*s by overriding any tendency they might have to withhold responses.

EXPERIMENT II

Method

Design and procedure.—The 20/0 group of Exp. I was replicated at both levels of interference in every particular except for MMFR instructions. During $MMFR_1$, *S*s were given 5 min. for recall, were asked to guess whenever they were not sure, and were told to use all the time in attempting recall. This was enough time for all *S*s. With $MMFR_2$, 20/0 payoff was introduced, but *S*s were also told that they must write some word on each of the 16 lines of the sheet. A 5-min. time limit was announced, but not enforced because it was too short for many *S*s, and $MMFR_2$ had, in effect, no time limit. Confidence rating was required exactly as in Exp. I.

Subjects.—There were 16 *S*s, 8 in each condition. All were male Hopkins undergraduate volunteers paid $1.25 in addition to whatever they earned in the experiment. They were naive to verbal learning experiments and were not informed of the payoff before $MMFR_2$ instructions.

Results

Response recall.—Table 2 presents both mean recall and total intrusions obtained in each cell of the experiment. Analysis of OL revealed a significant RI effect, $F\ (1, 14) = 62.95$, $p < .001$. Forced responding successfully lowered response criterion, for $MMFR_2$ recall was better overall than $MMFR_1$, $F\ (1, 14) = 5.52$, $p < .05$. The improvement in $MMFR_2$ was largely due to increased recall of "unlearned" OL, and the interaction of the RI effect (IL trials) with the MMFR variable was significant, $F\ (1, 4) = 7.97$, $p < .025$.

Confidence rating.—Confidence rating showed the same high accuracy and skewed distribution of intrusions and correct responses over the categories as was found in Exp. I. Also in common with Exp. I was the small difference between $MMFR_1$ and

$MMFR_2$ in mean confidence of correct responses. In $MMFR_1$, the mean confidence was 6.46, and it was 6.64 in $MMFR_2$. The correct responses elicited in $MMFR_2$ and not in $MMFR_1$ had, however, a mean confidence of 6.27, indicating some criterial effect on recall. Intrusions showed a large shift in confidence when forced responding was demanded. In $MMFR_1$ the mean for the 4 intrusions was 3.25, compared to the mean of 1.31 for the 66 intrusions of $MMFR_2$.

SDT analysis.—Intrusions and correct responses did not overlap sufficiently on the confidence scale for *d′* or beta to be determined. As estimate of these parameters for OL in Exp. II was, however, made by averaging the probabilities (unweighed) of correct responses and intrusions in each category of confidence with those obtained in the appropriate condition of Exp. I. For $MMFR_1$ of Exp. II, the resulting distributions yielded a *d′* estimate of 2.05 for OL after 2 trials of IL and one of 2.60 after 20 trials. The mean of the betas for the three most stringent categories of the averaged distribution was .64 after 2 trials and 1.18 after 20. The same averaging procedure for $MMFR_2$ resulted in an estimated *d′* of 2.97 and mean beta of approximately .10 for OL after 2 trials and a *d′* of 2.77 and beta of 1.05 after 20 trials. These estimates can be considered as only a rough indication that the SDT analysis of Exp. I was not contradicted by behavior in Exp. II.

DISCUSSION

The fact that *d′* for OL was unaffected by interference, but beta became more strict as interference increased indicates that a restriction of guessing and not a loss of OL strength (unlearning) is responsible for RI. If, alternatively, *d′* is considered to be a retention score that is corrected for guessing, then the constancy of *d′* indicates that no more OL was lost after 20 trials of interference than after 2 trials, and a conservative shift in guessing is still implicated in the production of measured RI. If, however, a strict criterion is responsible for the apparent forgetting, why was recall increased only about 30% in Exp. II when use of the lowest possible criterion was forced on *S*s? An answer to this question as well as to that of the mechanism responsible for the criterion shift is attempted in the following discussion of the role of generalized response competition in interference.

According to the hypothesis of generalized competition, IL reduces the set or bias to emit the class of responses constituting OL. The OL responses no longer form a coherent list, but must, as it were, be individually searched out in memory, and they are therefore no longer protected from competing items by a tendency to be called up together. These competing items could be IL responses, potential intrusions, and/or extraneous associations made available by weakened response bias. Such competition increases the number of possibilities among which *S* must search in recalling OL, and it is suggested here that this increase is responsible both for RI and for the interference-induced shift in criterion. An increase in the pool of items from which OL is selected will, by itself, reduce the amount recalled even though the competitors are all very low in strength and the strength of OL is unaltered. In the forced-choice psychophysical situation it is, in fact, possible to calculate the decrease in hit rate for targets with a constant *d′* as the number of distractors of zero mean strength and unit variance increase (Elliott, 1964). The MMFR situation differs in many important respects from the forced-choice situation analyzed by Elliott: The distribution of strengths of the distractors is not known, the distractors probably are not orthogonal, and the choices are not under experimental control. Only in the general sense that a decrement in OL recall after IL without loss of OL strength is predicted from generalized response competition can the analogy with the forced-choice situation be entertained.

The criterion, then, becomes more stringent after interference because of the increased pool of distractors among which *S* must select the OL responses. In order to avoid intruding the competitors in recall, *S* is more cautious in responding. This caution obtained also in Exp. II when responding was forced. The higher categories of confidence were still used as cautiously as in Exp. I, and the lowest category absorbed most of the competitors *S*s were forced to emit. The increase in recall obtained with forced responding is largely the result of forcing *S* to search his memory a bit longer and is only partly the result of lowered criterion. It seems likely that with the degree of interference afforded by 20 trials of IL, competition is so great that a small improvement in recall beyond the 30% obtained with forced responding would require a vastly extended search. It seems unlikely, however, that *S*s could be induced to search much longer than they did in the forced responding condition.

The main result of this experiment is comparable to Murdock's (1966) finding that *d′* remains constant over serial position (SP) in the short-term recall of paired associates, even though recall differs among the positions in accord with a SP effect. Criterion setting followed the recall pattern, being more strict for SPs for which recall is lower. A criticism of Murdock's study, derived from Bernbach's (1967) discussion of it, applies also to the present one. Arguing from a finite-state model of memory, Bernbach suggests that forgetting may move an item from a learned (R) state to a nonlearned (N) one. Thus *d′* would remain constant for a set of items after some are forgotten because this measure would simply reflect the unaltered strength of the remaining items. It is not clear, however, how this model can account for the change in criterion. Furthermore, the reduction in RI produced by forced responding cannot be explained unless there is some way for an item, without reinforcement, to return to State R from State N. Finally, there is the question of how interference causes an item to move from State R to State N. The difficulties with the extinction theory of unlearning apply to a finite-state model of memory just as well as to a continuous strength model, and the finite-state model does not seem to offer a mechanism fundamentally different from extinction.

The mechanism suggested here explains measured forgetting without any reliance whatever on extinction. The component of RI identified as unlearning is seen as the result of attempting to select the correct responses from a pool of distractors. These distractors appear as the response bias that allows OL to be isolated is diminished. RI results not from the loss of items, but from the failure of the search for them in a given context.

REFERENCES

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associates in transfer theory. *Journal of Experimental Psychology*, 1959, **58**, 97–105.

Bernbach, H. A. Decision processes in memory. *Psychological Review*, 1967, **74**, 462–480.

Egan, J. P. Recognition memory and the operating characteristic. (Tech. Rep. No. AFCRC TN-58-51) Bloomington, Ind.: Indiana University, Hearing and Communication Laboratory, 1958.

Elliott, P. B. Tables of *d′*. In J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

McGovern, J. B. Extinction of associations in four transfer paradigms. *Psychological Monographs*, 1964, **78** (16, Whole No. 593).

Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology*, 1940, **53**, 173–203.

Murdock, B. B., Jr. The criterion problem in short-term memory. *Journal of Experimental Psychology*, 1966, **72**, 317–324.

Newton, J. M., & Wickens, D. D. Retroactive inhibition as a function of the temporal position and interpolated learning. *Journal of Experimental Psychology*, 1956, **51**, 149–154.

Pollack, I., & Decker, L. R. Confidence ratings, message reception, and the receiver operating characteristic. *Journal of the Acoustical Society of America*, 1958, **30**, 286–292.

Postman, L. Mechanisms of interference in forgetting. Vice-presidential address presented at the meeting of the American Association for the Advancement of Science, New York, December 1967.

Postman, L., Stark, K., & Fraser, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 672–694.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. Decision processes in perception. *Psychological Review*, 1961, **68**, 301–340.

(Received January 16, 1969)

ournal of Experimental Psychology
969, Vol. 82, No. 2, 224–230

DECISION AND DECAY PROCESSES IN THE SHORT-TERM MEMORY FOR LENGTH[1]

KATHARINE I. BLICK[2]

University of Washington

The short-term memory (STM) for length was studied in a paired-comparisons (AX) paradigm. The six retention intervals (1, 2, 3, 5, 9, and 15 sec.) and the 19 possible lengths of the standard stimulus were randomly varied within blocks. Besides equality, there were 10 possible length differences between the standard and comparison stimuli, and these differences were presented equally often within blocks except that the probability of a difference was equal to that of no difference. Two well-trained *S*s were asked to make a same–different recognition judgment in Part 1 and both a same–different and a longer–shorter judgment in Part 2. There was a significant decline in performance over unfilled retention intervals in both parts, but superimposed on this was a tendency for recognition accuracy to increase from the 2-sec. to the 5-sec. delay. The performances for the two types of decisions in Part 2 were not closely related. Thus, evidence was found for both two decay processes and for two separable decision processes.

The retention of a single nonverbal stimulus has been extensively investigated in the classical psychophysical literature, but the effects of response biases were usually not separated from those due to the decline in discriminability as a function of time. This consideration is particularly critical in the investigation of the nature of the decision and decay processes. For example, Köhler (1923) hypothesized the time order error (TOE) as a description of short-term decay processes, but his and most of the later experimentation have used response biases as measures of the TOE. Thus, it has been unclear whether this effect appears in retention or only during the decision process. Ronken (1967) and Wickelgren (1969), however, measured the TOE with the d' from signal detectability theory, which is theoretically free of response biasing effects. Wickelgren reported finding no TOE effects in the retention of pitch, but Ronken found consistent and reliable retention asymmetries in the short-term memory (STM) for tone intensities. Although Ronken's *S*s under certain conditions showed better discrimination for one standard intensity than another, at no time did the discrimination of either the weaker or the stronger standards improve as a function of time, as Köhler's theory predicts.

Evidence supporting the hypothesis of a nonunitary decay of nonverbal items from STM is found in studies with adequate control of response biasing effects. The TOE that Ronken (1967) found changed sign as a function of delay, paralleling but not accounting for the decreasing accuracy of performance with increasing delays. Wickelgren (1969) felt that his results were more adequately described by a two-process theory of decay than by a single-process theory. Also, it would be difficult for a single-process theory to encompass the reminiscence effects seen with pitch recognition (Irwin, 1937), nonsense form recognition (Eriksen & Lappin, 1967), and visual location (Posner, 1967).

There is also evidence for a nonunitary decision process. Wickelgren (1969) stated that there appeared to be two separate phases in the higher–same–lower judgment of pitch in a paired-comparison paradigm. Evidence for two separable decision processes is also seen in Dallet, Wilcox, and D'Andrea's

---

[1] This research was supported by Grants 11-6068 and 11-6149 from the National Institute of Mental Health, United States Public Health Service. The research was completed in partial fulfillment of the requirements for the PhD degree at the University of Washington.

[2] Requests for reprints should be sent to Katharine I. Blick, who is now at Western Illinois University, Department of Psychology, Macomb, Illinois 61455.

(1968) study of the retention of a group of meaningful pictures by means of both a recognition and a type of recall test. They suggested that there may be information in the STM trace that is available for use in an *S*'s confidence ratings for a recognition decision, but which is not available for recall tests.

The present experiment used familiar unidimensional stimuli in a delayed comparisons task to analyze the decision and decay processes involved in nonverbal STM. Length was chosen as the memory dimension because perceived length is linear with real length (Teghtsoonian & Teghtsoonian, 1967) and perceived length differences are also linear with real length (Creelman, 1965).

To avoid the interpretive errors of earlier studies, several modifications of procedure were introduced. The *S*s were extensively trained on the verbal labeling of the length stimuli so that the accuracy of the labeling could be adequately estimated and so that practice effects would be minimized. Payoffs and feedback were used after every trial to maximize performance. Both the values of the standard stimuli and the delays were varied within blocks of trials, as were the absolute length differences between the standard and comparison stimuli. This was done to force *S*s to maintain consistent strategies or standards of performance over conditions.

In Part 1, *S*s were asked to make a same–different judgment on the comparison length relative to the standard on each trial. In Part 2, they were asked not only to make the same–different judgment, but also to judge the comparison as longer or shorter than the standard. These judgments were then separately analyzed, so that conclusions as to the nature of the decision process could be made from the effect of the second decision on the first. The interrelationships of the two decisions were also studied.

METHOD

Subjects.—The *S*s were two undergraduates, one male and one female. They both had normal or corrected to normal vision. They were volunteers who were paid for their services according to the schedules specified subsequently.

Stimuli and apparatus.—The stimuli were steel rods, .045 in. in diameter and ranging in length from $13\frac{9}{16}$ in. to 15 in. in 24 equal steps of $\frac{1}{16}$ in. each. Four rods of each length were used interchangeably.

The *S* was seated at a table covered with black felt. The backdrop was black felt. A plywood screen was interposed between *S* and the table. During a trial, *S* rested his head on a chinrest and looked monocularly through a 2.25-in. round shuttered hole. The hole was 13 in. above the table level, and the rods were placed 18.8 in. from the screen so that they were approximately 23 in. from *S*'s eye. The rods were manually positioned by using a strip of black plastic tape on the table as a guide; the tape ran parallel to the plywood screen, but neither began nor terminated in *S*'s visual field. The left-right position of the rods was varied irregularly.

Before each trial, a toggle switch was thrown to select the Industrial timer set at the predetermined delay interval for that trial. Thus, a separate timer was used for the presentation time and for each delay condition.

Training.—The 24 different rod lengths were ordinally assigned the names "one" to "twenty-four," corresponding, respectively, to the shortest and to the longest rod lengths. The *S*'s task was to learn the name of each rod length.

Each rod was presented for 1 sec. Immediately after *S* verbally attempted to guess the name of the rod, *E* stated its name. The intertrial interval averaged about 9 sec., being jointly determined by *S*'s latency and by the time necessary to position another rod.

The stimuli were randomly presented in blocks of 175. The *S*s were given 2–3 blocks a day for 3–5 days a week, with at least 15 min. between blocks. The *S*s were paid according to their naming accuracy and were told their earnings at the end of every block. The *S*s were trained until a series of 10 blocks showed only random variation in naming accuracy.

Short-term memory task, Part 1.—The first or standard stimulus rod was presented for .75 sec. After a variable delay, the second or comparison rod was presented. The comparison rod remained visible until *S* responded with a verbal "same" or "different." The *S* then pushed a switch to close the shutter.

There were six possible delays between the first and second rods: 1, 2, 3, 5, 9, and 15 sec. The delays were randomly presented within each block of trials, except that no one delay could occur more than three times in a row. The intertrial interval was also variable, but averaged about 12 sec. per trial. The *S*s were instructed not to concentrate on their image of the standard rod during the delay interval because in pretests it was found that such concentration increased the variability of the data without affecting the mean accuracy.

There were 11 possible length differences between the comparison and standard stimuli: 0, $\pm\frac{1}{16}$, $\pm\frac{2}{16}$, $\pm\frac{3}{16}$, $\pm\frac{4}{16}$, or $\pm\frac{5}{16}$ in. Smaller and larger comparisons could be made equally likely for each standard length because only the rod lengths numbered 6–19 comprised the set of standard stimuli. The comparison set included all 24 lengths. "Same" and "different" presentations were equally likely, and the length differences were presented randomly within a given block. Thus, there were $10 \times 6 = 60$ possible combinations of "different" and the six delays, and 100 trials per *S* were run for each combination. Since 1,000 "same" trials were also run at each delay, there was a total of 12,000 trials per *S*. Several hundred practice trials were also given to each *S*.

There were 148 trials in a block, with a 5-min. rest halfway through the block. There were 3 warm-up trials at the beginning of each block and after the break in the middle. The entire block took about 50 min. to complete. The *S*s were run 2–3 blocks a day for 3–5 days a week.

The *S*s earned 7.5¢ for each correct decision, and 7.5¢ was subtracted for each incorrect judgment. As soon as possible after each block, *S* was informed of his earnings. The *S*s were repeatedly encouraged to try to better their performance.

Short-term memory task, Part 2.—In Part 2, *S*s made a longer–shorter judgment as well as a same–different judgment, so that a typical trial response would be "different, longer." Because *S*s were to make a longer–shorter judgment even if they first responded "same," they were told: "Since you are frequently wrong when you respond 'same,' tell me what the direction of the difference would be if, in fact, there was a difference." After two blocks of practice, *S*s had no real trouble in making the combined judgment.

Three delays were used in this part of the experiment: 1, 2, and 5 sec. There were 75 trials for each of the $10 \times 3 = 30$ different possible combinations of "different" and the three delays. There were also 750 "same" trials, making a total of 4,500 trials for each *S*.

The same–different judgments in Part 2 had the same payoff as in Part 1; the payoff for the longer–shorter judgment was ±2¢ and was computed only for the trials where there actually was a difference present. All other conditions were the same as in Part 1.

RESULTS

Both *S*s had a bias towards saying "same," and this bias tended to decrease regularly as the time delay was increased. Because of this bias, the *d′* of signal detectability theory was used in the data analysis. The *d′*s were taken from Swets' (1964) table. For Part 1 and for the same–different decision of Part 2, the probabilities used for entering the table were the probability of a "different" response given that there was, in fact, a length difference between standard and comparison and the probability of a "different" response given no difference. The same conditional probability of a "different" response given no difference was used for the 10 possible standard-comparison differences at a given delay. The *d′* of the longer–shorter decision was determined by means of the probability of a "longer" response given that the comparison stimulus was longer than the standard and the probability of a "longer" response given a shorter comparison. These methods for determining *d′* assume that the longer–shorter and the same–different decisions are made separately; evidence for this assumption is discussed later. No inferences should be made concerning the relative accuracy of the two decisions. Since accuracy in the training task did not vary with rod length, all the standards are treated as equivalent in the following analyses. The tests of significance (all two-tailed unless otherwise specified) and estimates of variance come from Gourevitch and Galanter (1967).

Figures 1 and 2 present *d′* as a function of delay for A. S. (male) and S. S. (female), respectively. The data from both decisions and both parts of the experiment are included. Since the *d′*s were averaged over the standard-comparison differences for these figures, each point from Part 1 represents 2,000 observations, each point from the same–different decision of Part 2 represents 1,500 observations, and each point for the longer–shorter decision represents 750 observations. The time delay curves for most of the comparison-standard differences taken individually tend to be of the same form as the average, although the position of the second peak in Part 1 varies irregularly. Specifically, although *S*'s performance with the longer comparisons differs from that with the shorter comparisons (as is discussed later), the form of the retention function is the same for both the longer and shorter comparison lengths. The standard deviation varies from .05 to .07 for all same–different points and is approximately .09 for the longer–shorter points.

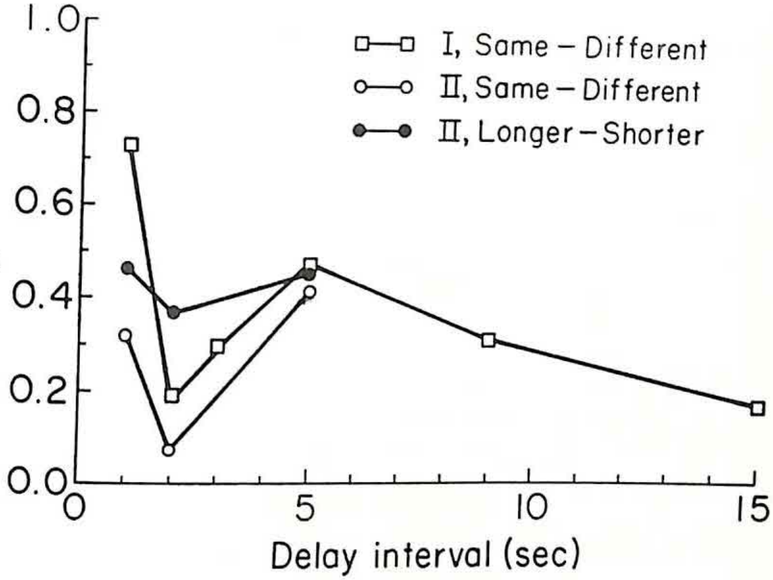


FIG. 1. *d′* as a function of delay for A. S. from both Parts 1 and 2 and for both decisions.

For A. S. in Part 2, the *d′* at 5 sec. is significantly different from those at 2 and 15 sec. ($p < .005$) and from those at 3 and 9 sec. ($p < .05$). The *d′* at 1 sec. significantly differs from all others ($p < .005$). In A. S.'s same–different performance in Part 2, the *d′* at the 2-sec. delay significantly differs from that at 1 sec. ($p < .01$) and from that at 5 sec. ($p < .005$). A. S.'s Part 2 same–different performance is significantly different from that of Part 1 at the 1-sec. delay ($p < .005$), but not at 2 or 5 sec. There are no significant differences among the longer–shorter *d′*s.

For S. S. in Part 1, the *d′* at 1 sec. is significantly different from all others ($p < .005$), but there are no other significant differences. For this *S* in Part 2, the same–different *d′* at 2 sec. was also significantly different from that at 1 ($p < .05$) and at 5

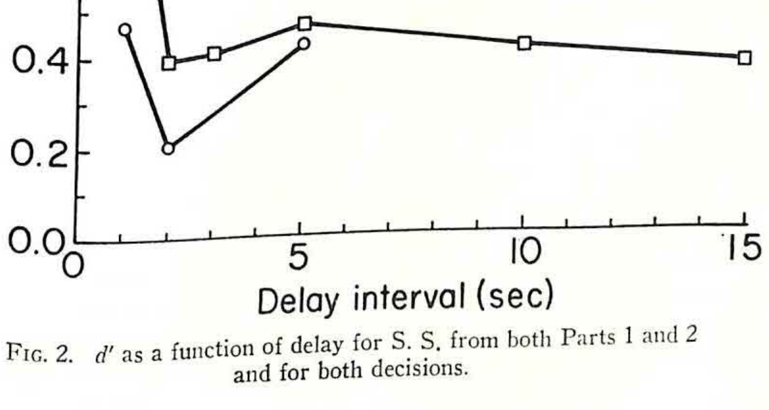


FIG. 2. *d′* as a function of delay for S. S. from both Parts 1 and 2 and for both decisions.

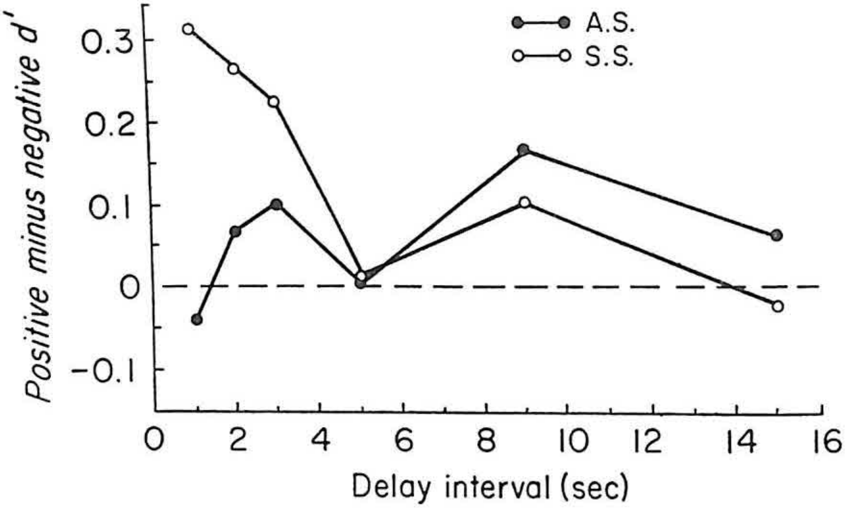


FIG. 3. The TOE functions for both *S*s as a function of the retention interval.

sec. ($p < .01$). S. S.'s same–different performance is depressed in Part 2 relative to Part 1. This depression is significant at both the 1- and 2-sec. delays ($p < .005$ and .05, respectively). Again, there are no significant differences among the longer–shorter *d*'s.

Figure 3 indicates TOE effects. The TOE is defined here as the differential sensitivity to positive (comparison larger than standard) as opposed to negative (smaller comparisons) differences. The data points in Fig. 3 represent the mean *d'* of performances, with shorter comparison lengths subtracted from the mean *d'* of the longer comparisons. These mean differences are graphed against delay for each *S*. Each point thus represents 500 shorter and 500 longer observations. The only significant differences between positive and negative *d*'s were for S. S. at the 1-, 2-, and 3-sec. delays ($p < .005$, .01, and .05, respectively). There were no significant TOE effects at any delay for either *S* in Part 2, although the positive *d*'s were larger than the negative *d*'s for both *S*s at all three delays.

The larger and smaller comparison lengths of equal differences were combined so that *d'* could be plotted against the comparison–standard length differences in sixteenths of an inch. Because there was no significant Delays × Comparison-Standard Differences interaction, the data points were also averaged over the time delays. Inspection of the resulting figures provided no reason to reject an assumption of a linear decline in performance as a function of decreasing standard-comparison differences.

The probability of a "same" response, given each possible length difference between standard and comparison, was plotted for each time delay, for each part, and for each *S*. The resulting functions were variable and permitted only the conclusion that discrimination tended to flatten with less accurate performances, due mainly to the decreasing probability of a correct "same" response.

One method of evaluating the relationship between the two decisions is to examine the probability of a correct longer–shorter response given that *S*s first incorrectly called the stimulus "same." If the decisions are made simultaneously, then the "second guess" should be random and would contain no information not present in the "same" decision. The probabilities of a correct second guess for A. S. at 1, 2, and 5 sec. are .509, .503, and .538, respectively. The corresponding values for S. S. are .550, .507, and .582. The chi-square (one-tailed with 1 *df*) was not significant at any one time or with all delay times averaged together for A. S. S. S.'s chi-square values were significant at 1 sec. ($p < .05$), at 5 sec. ($p <$

.005), and for all delay times taken together ($p < .005$).

Another way of evaluating the relationship between these decisions is to determine what percentage of the correct longer–shorter responses are preceded by a correct response of "different." This conditional probability was .611 for A. S. and .688 for S. S.

Discussion

Both *S*s in the present experiment showed a significant decline in accuracy as a function of an unfilled time delay. This decline was probably not due to the decay of a verbal label because *S*'s labeling was asymptotic before the STM test and the retention of a single number should not decline over an unfilled interval. Furthermore, *S*s' memory performances were more accurate than their labeling. Thus, there was probably a large nonverbal component involved in the present study of length retention. However, Posner and Konick (1966) found no decay over unfilled delay intervals for the STM of the position of a circle on a line. They did find a decay over unfilled intervals for their kinesthetic location task, which was also more difficult for their *S*s than was the visual task. Thus, difficulty may be a relevant variable in nonverbal STM.

Both Ronken (1967) and Wickelgren (1969) found that an exponential function was an adequate, if not the best possible, description of their forgetting curves; this is not true in the present experiment. Because practice effects were eliminated, the consistent improvement in performance from the 2- to 5-sec. delay cannot be due to a more rapid improvement of comparison ability at one delay than at another. This improvement is similar to the effects seen by previous investigators with verbal (e.g., Keppel & Underwood, 1967) and with nonverbal material (Eriksen & Lappin, 1967; Irwin, 1937; Posner, 1967). However, while the performance at 5 sec. is superior to that at 2 sec. in every one of the six possible comparisons, the superiority is significant in only three. If there is a reminiscence effect, it is relatively weak, but a unitary theory of STM will have trouble with even a weak effect.

A unitary decay theory will also find it difficult to incorporate the TOE effects found in the present experiment and in Ronken's (1967). Since performances with neither the longer nor the shorter comparison lengths improved with increasing delays, the classical TOE theory is erroneous. However, because the TOE depends on a performance and not a response bias measure in the present and in Ronken's experiments, TOE effects might have to be theoretically considered. However, the conditions necessary for the appearance of a TOE effect are unclear since neither Wickelgren (1969) nor Posner (1967) found any evidence for a TOE in pitch or in length judgments.

Generalization effects appear to be quite similar for length and for pitch (Wickelgren, 1969). In both cases, the gradients flatten with increasing delays (or with decreasing accuracy).

Adding a longer–shorter decision to the same–different decision already required depresses the accuracy of the latter. Increasing the number of decisions to be made might increase the rate of decay, possibly by increasing the amount of information that the trace must carry to make both decisions possible. In verbal tasks, increasing the amount of information can increase the rate of forgetting (e.g., Merikle, 1968). Increasing the number of decisions might also increase the decision time and thereby increase the delay interval. In partial support of this is the fact that most of the decrement from Part 1 to Part 2 was due to a decreased probability of a correct "same" response, which was also the main effect of increasing the delay from 1 to 2 sec. in both Parts 1 and 2.

In either case, it seems as if the decision process takes up "space" in STM, so that accuracy is decreased by additions to the decision process. Further indications that the decision process has dimensional properties come from the evidence for the separation of the two decisions. Success in the same–different decision does not adequately predict success in the longer–shorter decision, and there is significant information in the second guesses for S. S. Also, *S*s reported that they made the decisions separately. This corresponds to the evidence that Wickelgren (1969) and Dallet, Wilcox, and D'Andrea (1968) found for two separable decision processes.

Taken together, the present experiment and those cited previously provide evidence for multiple processes in STM for nonverbal material. The reminiscence and TOE effects cannot be handled by a single-process decay theory with any degree of simplicity. The performances in the present experiment at 1 and 2 sec. differ both quantitatively and qualitatively from those at 5, 9, and 15 sec. S. S. and A. S. had similar TOE functions from the 5- to 15-sec. delay, but not from the 1- to 2-sec. delay. The detrimental effects of the longer–shorter decision are most marked at the 1- and 2-sec. delays and are almost absent by the 5-sec. delay. However, since both types of responses have similar decay functions in Part 2, both decisions are probably subject to both types of decay processes. These results show that neither the short-term decay nor the recognition process are unitary phenomena when familiar nonverbal stimuli, differing on one dimension, are presented singly to STM.

REFERENCES

Creelman, C. D. Discriminability and scaling in linear extent. *Journal of Experimental Psychology,* 1965, **70**, 192–200.

Dallet, K., Wilcox, S. G., & D'Andrea, L. Picture memory experiments. *Journal of Experimental Psychology,* 1968, **76**, 312–320.

Eriksen, C. W., & Lappin, J. S. Selective attention and very short-term recognition memory for nonsense forms. *Journal of Experimental Psychology,* 1967, **73**, 358–364.

Gourevitch, V., & Galanter, E. A significance test for one parameter isosensitivity functions. *Psychometrika,* 1967, **32**, 25–33.

Irwin, C. C. A study of differential pitch sensitivity relative to auditory theory. *Journal of Experimental Psychology,* 1937, **21**, 642–652.

Keppel, G., & Underwood, B. J. Reminiscence in the short-term retention of paired-associate lists. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 375–382.

Köhler, W. The problem of form in perception. *British Journal of Psychology,* 1923, **14**, 262–268.

Merikle, P. M. Unit size and interpolated-task difficulty as determinants of short-term retention. *Journal of Experimental Psychology,* 1968, **77**, 370–375.

Posner, M. I. Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology,* 1967, **75**, 103–107.

Posner, M. I., & Konick, A. F. Short-term retention of visual and kinesthetic information. *Organizational Behavior and Human Performance,* 1966, **1**, 71–86.

Ronken, D. A. The constant error in amplitude discrimination and the inter-stimulus interval: A re-examination. Unpublished doctoral dissertation, University of Washington, 1967.

Swets, J. A. *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Teghtsoonian, M., & Teghtsoonian, R. Seen and felt length. *Psychonomic Science,* 1967, **3**, 465–466.

Wickelgren, W. A. Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology,* 1969, **6**, 13–61.

(Received February 13, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 231–237

EFFECTS OF EXTENDED PRACTICE ON HIGH-SPEED SCANNING[1]

DAVID BURROWS AND BENNET B. MURDOCK, JR.[2]

University of Toronto

High-speed scanning of short-term memory was studied with a technique similar to that of Sternberg. The *S*s were required to decide if any one of a set of memorized targets (*t*) was present in a set of briefly presented visual items (*d*), then make a binary response based on this decision. Latency of this response was sensitive to increases both in *t* and *d*. It was hypothesized that an underlying serial-serial search process was involved, with each target item being compared to each display item in a one-to-one fashion. Extended practice made this comparison process more rapid, but did not alter its serial-serial nature.

One of the most important problems of short-term memory (STM) is that of retrieval; how is stored information found when it is required for output? There is some recent support for models of retrieval that hypothesize a search or scan of the contents of a memory store when a particular item of information from such a store is required (Atkinson & Shiffrin, 1967; Sternberg, 1966; Yntema & Trask, 1963).

Sternberg (1966) found that when *S*s were asked to decide whether a single visually shown digit was contained in a memorized sequence of digits, mean reaction time (RT) for the decision increased with the number of digits in the sequence. This finding suggests that the latency of the decision reflects an internal serial comparison process in which the displayed digit is compared with each of the memorized digits, one at a time. Furthermore, the slope of the increase in RT with memorized sequence length was identical for both positive and negative responses. Since *S* must search through all digits in the memorized sequence before making a negative response, this equality of slopes suggests that *S* searches through the entire sequence on positive trials also. Sternberg calls this process serial exhaustive scanning.

An alternative search method might be a self-terminating search in which a response would be made as soon as a "match" is found or, if no match is found, after the entire set has been searched. For a set of s items with a "critical" or matching item randomly located in the set, a positive response can be made after searching through an average of $(s + 1)/2$ items. If there is no critical item, a negative response is made after searching through s items. Increases in set size s will have twice as much effect on negative responses as on positive responses, and the slopes, as s increases, should have a ratio of 2:1 for negative and positive responses, respectively.

Nickerson (1966) extended Sternberg's (1966) finding to the case where the number of visually presented items, as well as the number of memory items, is varied and a positive response is required if the two sets have any elements in common. Nickerson found that RT increased with increases in the size of either set. Nickerson also found that prolonged practice could attenuate but not eliminate these effects.

---

[1] This study was an MA thesis by the first author under the direction of the second author. The research was supported by Grants 164 from the Ontario Mental Health Foundation, APA 146 from the National Research Council of Canada, and GB 4545 from the National Science Foundation. The authors would like to thank members of the Ebbinghaus Empire for their comments and Ron Okada for help with computer programming and data analysis.

Subsequent to the preparation of this paper, the authors became aware of similar results reported by S. Sternberg, "Scanning a Persisting Visual Image versus a Memorized List," a paper read at the meeting of the Eastern Psychological Association, Boston, April 1967.

[2] Requests for reprints should be sent to Bennet B. Murdock, Jr., Department of Psychology, University of Toronto, Toronto 5, Canada.

Contradictory results have been reported by Neisser and his associates (Neisser, 1963; Neisser, Novick, & Lazar, 1963) for experiments in which *S* searched through rows of visually presented characters until finding any of one or more memorized "target" items. Neisser found that with extended practice on the same set of targets, up to 10 different items can be searched for as rapidly as a single item. Neisser (1967) has argued for a comparison process that can become so efficient with practice that multiple memory targets can be tested for in parallel when searching a visual display. The Sternberg (1966) and Nickerson (1966) experiments, which found evidence for serial testing of memory items, differed from Neisser's in that the former tested *S*s with a discrete trials procedure and penalized them for mistakes, whereas Neisser's *S*s scanned through a continuous display and were encouraged to go as rapidly as possible with less explicit emphasis on accuracy. In addition, Neisser used "fixed" memory targets (same memorized sequence on each trial), while Nickerson used varied sets (different sequence on each trial). Sternberg has compared these two conditions under limited amounts of practice finding no difference, but the possibility exists that differences could turn up with greater amounts of practice.

The purposes of the present experiment were: (*a*) to attempt to resolve the issue of serial vs. parallel tests for items in memory using a discrete trials procedure, with errors controlled and extended practice, for both a "fixed" target and a "varied" target procedure; and (*b*) to examine in detail the processes involved in comparing several memory items with several visual items while looking for a match. It is possible that a set of visual items can be scanned in either a serial or parallel fashion and be self-terminating or exhaustive and that these procedures can change with practice, just as all of these possibilities exist for memory scan.

Parallel processing is defined here as the capacity to handle multiple inputs as efficiently as single inputs. Invariance of RT with increasing number of items to be handled will be taken as evidence for parallel processing. A serial process is defined in the present paper as any process that produces an increase in RT with increased load.

It should be noted that the present experiment cannot eliminate parallel-processing theories which postulate simultaneous comparisons with *rate* of comparison dependent on number of comparisons. The present issue is whether additional processing load does or does not affect processing time.

METHOD

Subjects.—The *S*s were four female first-year psychology students from the University of Toronto. Each *S* was paid $1.50 per session for participation in the experiment and was not informed of the purpose of the experiment until completing all of the sessions.

Apparatus.—Stimuli were typed on 2 × 2 in. radio-mat slides and presented by a Kodak carousel 600 projector equipped with a 500-w. bulb on an opaque screen approximately 12 ft. from *S*. Digits were approximately 3.2° high, and the widest "row" of digits was 13° wide. Exposure duration was controlled by a Hunter interval timer. The luminance of the display was 90 ftl., as measured by a Spectra brightness spotmeter. Preexposure and postexposure fields were dark. Reaction time was defined as the interval between the beginning of the stimulus exposure and the initiation of *S*'s response, which consisted of depressing one of two push-button switches mounted in a horizontal row on a board resting in front of *S*. The RTs were measured to the nearest .01 sec. by a Standard Electric timer. The *S*s were instructed to use the first finger of their left and right hands to control the buttons.

Procedure.—On each trial, *S* was read a series of 1, 2, or 3 digits at a 40 digits/min rate, was given a "ready" signal, and was presented with a slide containing a horizontal array of 3, 4, 5, or 6 digits for a 300-msec. duration. The slide was presented 2 sec. after the last digit was spoken. The *S*s were told to fixate at a point representing the center of the exposure field, decide if *any* of the spoken "target" digits had been presented in the visual "display," and press a no or yes key accordingly. The *S*s were informed that one and only one target digit would appear in the display on exactly 50% of the trials and no target digits would be in the display on the remaining trials. The intertrial interval was 6 sec.

For each experimental session, three target lengths (1, 2, or 3 digits) were varied orthogonally with four display lengths (3, 4, 5, or 6 digits) both for positive and for negative instances. For each Yes–No × Target Length × Display Length combination there were eight replications, giving 2 × 8 × 3 × 4, or 192 trials per session. Each *S* served in 15 consecutive daily sessions (weekends excluded). Day 1 was considered practice for each *S* and was excluded from the analysis.

The *S*s were divided into two experimental groups of two each. For the varied-target group, the target digits for each trial were randomly selected without replacement from the set 0–9. For the fixed-target group, the same digit(s) were used for each target length condition throughout the entire 15 sessions. Thus an *S* in the fixed-target condition would be presented with the same 1, 2, or 3 digits whenever the target length was 1, 2, or 3. Furthermore, for each *S* in the fixed-target group the multiple-digit targets were built up from simpler targets in the following manner: If "8" made up the one-digit target for a fixed-target *S*, the two-digit target was "8" and some other digit and the three-digit target consisted of the two-digit target plus one other digit.

The 64 trials involving the same target length conditions were run in a consecutive block for all *S*s in each session. The order in which each target length block was run was varied randomly from session to session. Within each target length block, display length and yes–no were ordered randomly.

The display digits were selected randomly from the set 0–9 on each trial, subject to the constraints that for each positive trial, one display digit must also have been in the target and that for each negative trial, the display digits and target digits have no digits in common. For positive trials, the common digit appeared in each possible serial position of the target with equal probability and in each possible serial position of the display with equal probability. Assignment of yes and no buttons was counterbalanced across handedness, and before every session each *S* was cautioned to be as accurate as possible.

The design of the experiment was thus a seven-way factorial, 8 × 2 × 4 × 3 × 14 × 2 × 2, with the following factors: replications, yes–no, display length, target length, days, *S*s, and fixed or variable target. The *S*s were nested under fixed or variable target, and all other factors were completely crossed.

RESULTS

Since the main interest is in search processes under conditions of errorless performance, scores for all trials on which errors were made were eliminated and replaced by mean latency for the remaining replications under the same treatment conditions. No *S* made errors on more than 4% of the total trials, and group error rates ranged from 2% for the condition in which target length was 1 and display length was 3 to 7% for the condition in which target length was 3 and display length was 6.

The latencies were analyzed using a partially hierarchical classification (Winer, 1962). There was no effect for replications

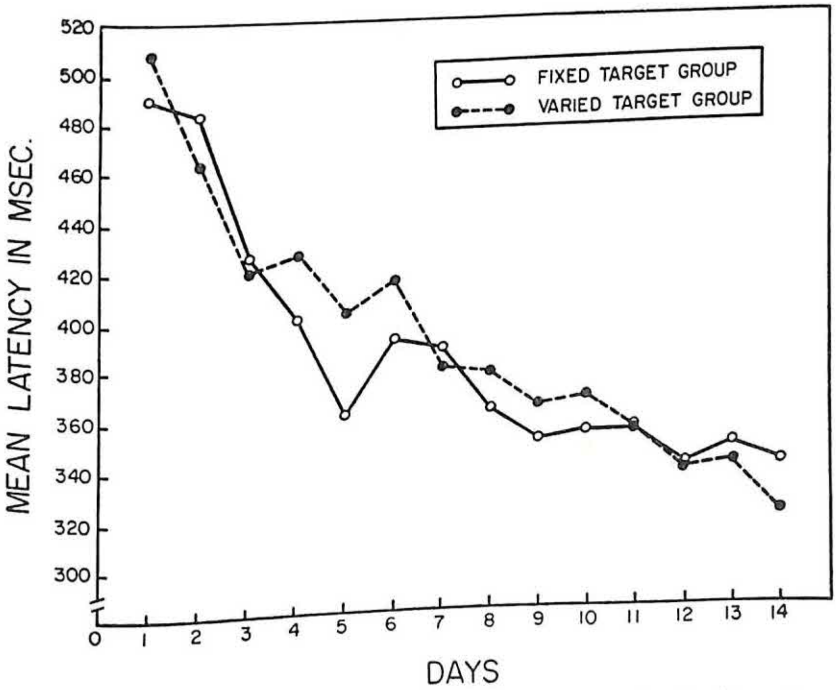


FIG. 1. Mean latency of response as a function of days for fixed-target and varied-target groups.

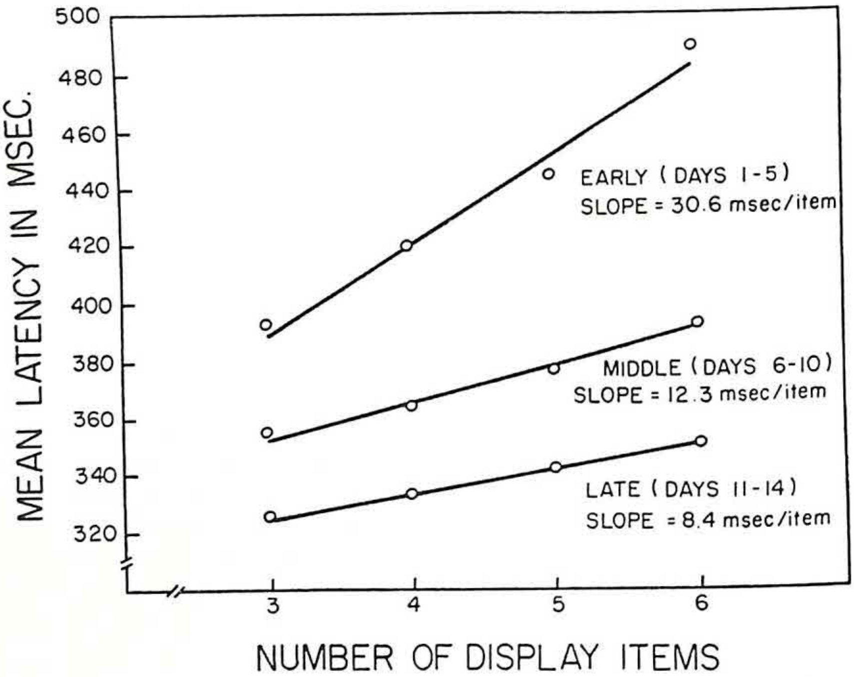


FIG. 2. Mean latency of response as a function of display length for early, middle, and late stages of practice.

($F < 1$) or any interactions involving replications. Therefore, data for all replications were combined, and the remaining analyses were based on means over replications.

Effects of practice.—Figure 1 shows mean latency for all trials on each day for both the fixed-target and variable-target groups. Latency decreased with practice, as indicated by the days effect, $F\ (13, 26) = 17.9$, $p < .01$.

This improvement is invariant over type of target: fixed vs. varied target was not significant, $F < 1$, and did not interact with days or any other factors.

The absolute values of these latencies are somewhat lower than those for similar experiments reported in the literature. This appears to be due in part to some constant error of measurement introduced into the origin. Since the changes in latency over different conditions and therefore the *relative* values of the latencies are of interest, this error of origin does not seem to be important.

Figure 2 shows latency as a function of display lengths for early (Days 1–5), middle (Days 6–10), and late (Days 11–14) stages of practice, with fixed- and varied-target groups combined. The straight lines were fitted by least squares, with linear regression accounting for 93% of the variance.

Latency is a linear function of display length for all stages of practice, as shown by Fig. 2 and the display length effect, $F\ (3, 6) = 10.37$, $p < .01$. The overall linear trend was significant, $F\ (1, 6) = 28.8$, $p < .01$, while the overall deviation from linearity was not significant, $F\ (2, 6) = 2.17$.

The magnitude of this effect declines with practice: the slopes are decreasing and the Days × Display Length interaction is significant, $F\ (39, 78) = 5.4$, $p < .01$. The practice effect seems to be due to this decrease in slope: extrapolations of the three fitted lines show no systematic change in zero intercepts, and the only other interaction involving days is Yes–No × Days, $F\ (13, 26) = 4.02$, $p < .01$, which is a consequence of the decreasing display effect. Yes–no and display length are interdependent in a manner indicated later.

It appears that *S*s serially scan the display, scanning *faster* as they become practiced. Of particular importance is the absence of the Days × Target Length interaction, F (26, 52) = 1.56, $p < .10$; although target length is significant, F (2, 4) = 41.6, $p < .01$, practice does not attenuate the effects of memory load.

Target length and display length.—Latency is sensitive to both target length and display length, as indicated by the main effects for these factors. Figure 3 shows mean latency of response as a function of display length for each target length and yes–no condition. The fitted lines were obtained by least squares with linear regression accounting for 92% of the variance.

For all cases, latency increases with display length. For each target length the slope of the function is twice as steep for the negative trials as it is for positive trials. This finding and the Display × Yes–No interaction, F (3, 6) = 7.26, $p < .05$, suggest a serial, self-terminating scan of the display.

The effect of increasing display lengths (d) seems to be enhanced by increases in target length (t), as evidenced by the increasing slopes for $t = 1, 2, 3$. This implies a serial search of items in memory.

This sensitivity to both target and display length suggests a basic process of one-by-one comparisons between display and memory items, although this conclusion must be tempered by the nonsignificant Display Length × Target Length interaction, F (6, 12) = 2.87, $p < .10$.

The negative–positive differences imply that the comparisons continue until either a match is found or all possible comparisons have been made. There are at least two different orders in which items from target and display sequences might be selected for comparisons: (a) Display continuous. The first display item is compared in turn to the first target item, second target item (if there is one), etc., through the target. Then the process is repeated with the second display item. The process ends either when a match is made or all comparisons have been exhausted. It is "display continuous" because each display item is compared to all target items without interruptions for comparisons involving other display items. (b) Memory continuous. The first memory item is compared to all display items, then the second memory item is compared to the display, etc.

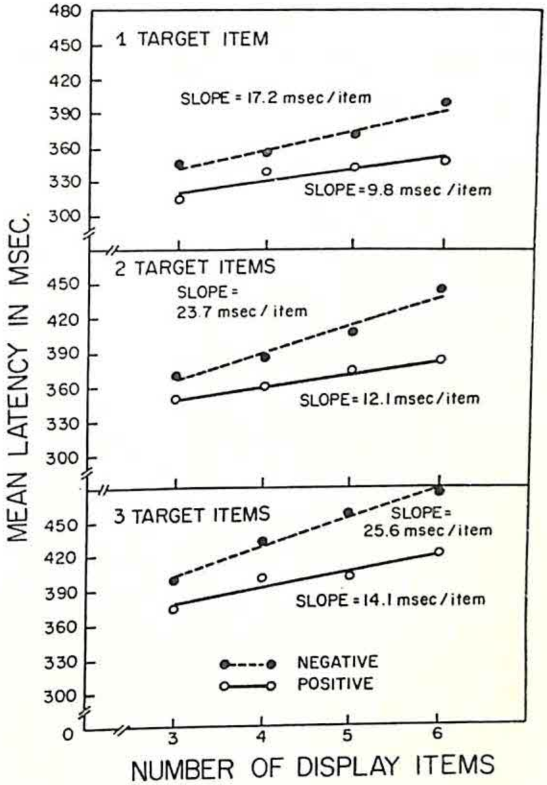


FIG. 3. Mean latency of response as a function of display length for each target length and positive–negative condition.

If all such comparisons proceed in a forward order (i.e., from presented first to presented last in memory and left to right in display), then for positive responses it should be possible to distinguish between the two hypotheses about scanning order on the basis of the effect of the position of the "critical" item in both the target and the display.

If Hypothesis a is true, then RTs would tend to be shorter for positive trials where the critical item is first in both target and display, next shortest when the critical item is second in target and first in display (assuming at least two targets), followed by third in target, first in display, first in target, second in display, etc. Hypothesis b predicts a different rank order of latencies, e.g., second fastest responses to critical items *first* in target and *second* in display. Table 1 shows Spearman rank-order correlations be-

TABLE 1

Spearman Correlation Coefficients between Rank-Order Latencies for Display-Continuous and Target-Continuous Hypotheses and Obtained Rank Orders

Number of items in display (d)	Number of items in target (t)	r_s	
		Display-Continuous Hypothesis a	Memory-Continuous Hypothesis b
3	2	−.030	.486
3	3	.600*	.600*
4	2	.214	.690*
4	3	.504	.346
5	2	.327	.269
5	3	.469*	.612**
6	2	.492	.567*
6	3	.228	.610**

* $p < .05$.
** $p < .01$.

tween rank orders of latencies predicted by each hypothesis and the obtained rank orders for the cases where $t = 2$ and 3 and $d = 3$, 4, 5, and 6. Data are for the varied-set *S*s only since target position has little meaning for the fixed-set *S*s. The results, which are representative of both *S*s, support the memory-continuous hypothesis.

Discussion

The results seem to favor a serial-serial scanning mechanism that is self-terminating in both the display and memory store. The first item in the target is compared to the first item in the display. Then it is compared to the second item in the display, to the third item in the display, etc., through the display. Then the second item in the target is compared to each item in the display, one at a time. If no match is found after all possible comparisons have been made, a "no" decision is made. If any match is found, scanning is terminated and a "yes" decision is made. If there are d items in the display and t items in the target, *S* must make $t \times d$ comparisons on the negative trials. On positive trials, the number of comparisons that *S* must make before arriving at a decision is determined by the location of the critical item in both target and display. Since the location of the critical item in both target and display is random, each of the possible total number of comparisons, 1, 2, 3, . . ., $t \times d$, is equally likely and the *mean* number of comparisons is $[(t \times d) + 1]/2$.

With t constant, increases in d will have twice as great an effect on negative responses as on positive responses. With d constant, increases in t should also affect the negative functions twice as much. The observed 2:1 ratio of negative to positive slopes at all values of t agrees with both requirements.

Evidence against a serial-serial model is provided by the nonsignificance of the Target × Display interaction. In addition, a serial model would predict slopes in the ratio 1:2:3 for $t = 1, 2, 3$. The observed changes in slope are not this great. One possible explanation for this is that the process is, in fact, serial in that comparisons are not initiated simultaneously, but that different comparisons may overlap in time; some comparisons may be initiated before others are finished. This would account for the sensitivity of latency to number of items in memory (which a strictly parallel model as defined here cannot do) and for the fact that slope changes are not as great as expected.

Exhaustiveness and serial position effects.—In addition to the present experiment, several studies have reported evidence contrary to Sternberg's (1966) exhaustive scan model. Norman (1966), Morin, DeRosa, and Stultz (1967), and Corballis (1967) all find RTs for positive responses to be affected by the serial position in the memory set of the matching item; if a set is scanned exhaustively, the position of the critical item should have no effect. Nickerson (1966) also found evidence against *complete* exhaustiveness for multiitem displays. Why scanning of memory appears to be exhaustive under some circumstances but self-terminating under others is a topic for further research.

The results obtained by Morin et al. (1967) and Corballis (1967) showed primacy and recency effects (shorter latencies for memory items presented first and last than for middle items), whereas the hypothesized scanning order suggested by the data from the present experiment predicts primacy only. This discrepancy may be justifiable in terms of procedural differences; the other studies used shorter warning intervals (less than .5 sec.) at which some sensory trace of the last items may still be available (cf. Posner & Keele, 1967; Sperling, 1960). The sensory trace may be checked first before the memory store itself is searched. In addition, order of scanning may be affected by strategy considerations related to the number in memory to be scanned. When STM is overloaded with items, as in Norman's (1966) study using 15-item lists, a strategy of scanning the most recently presented items first, while their sensory trace is still available, may maximize the number of items that can be checked before they are lost from memory. In the present study, the number of memory items is small enough so that all items could be scanned in a forward order without any loss from memory. Scanning order may indeed depend on parameters such as set size and warning interval.

Serial processing, parallel processing, and the effects of practice.—The results agree with several findings that processing time depends on the number of items in memory when a varied set is used (Kaplan & Carvellas, 1965; Kaplan, Carvellas, & Metlay, 1966; Nickerson, 1966; Sternberg, 1966) and with Nickerson's finding that practice does not eliminate this effect. The results corroborate Sternberg's finding that the memory load effect does not depend on whether a varied- or fixed-set procedure is used. More importantly, this last finding holds up under very extensive practice. This stands in almost direct opposition to the results reported by Neisser et al. (1963), who found that several targets could be searched for simultaneously. This discrepancy might be accounted for by the strict demands placed on *S*'s performance in the present experiment. It is possible that Neisser's technique of placing emphasis on speed rather than accuracy and using somewhat loose controls may have resulted in a failure to find effects of target length which, though small, did in fact exist. What may appear to be a parallel process may in fact be an extremely rapid serial process.

Extraction of information from STM and from a visual display appear to be similar in that both involve some sort of serial scan, given the definition of serial as an increase in RT with increase in load. The two processes do, however, appear to be affected somewhat differently by practice. Scanning of STM appears to be unique in that as long as several items are held in memory, each one is retrieved or extracted one at a time. The crucial factor in search load seems to be the number of items in immediate memory, regardless of how well overlearned each item is. As long as an item exists in immediate memory as a possible target, it must be scanned, at the expense of some unit of time. The fact that such an item may also be represented in a long-term memory store does not eliminate this requirement.

REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. (Tech. Rep. No. 110) Stanford, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1967.

CORBALLIS, M. C. Serial order in recognition and recall. *Journal of Experimental Psychology,* 1967, **74**, 99–105.

KAPLAN, I. T., & CARVELLAS, T. Scanning for multiple targets. *Perceptual and Motor Skills,* 1965, **21**, 239–243.

KAPLAN, I. T., CARVELLAS, T., & METLAY, W. Visual search and immediate memory. *Journal of Experimental Psychology,* 1966, **71**, 488–493.

MORIN, R. E., DEROSA, D. V., & STULTZ, V. Recognition memory and reaction time. *Acta Psychologica,* 1967, **27**, 298–305.

NEISSER, U. Decision time without reaction-time: Experiments in visual scanning. *American Journal of Psychology,* 1963, **76**, 376–385.

NEISSER, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

NEISSER, U., NOVICK, R., & LAZAR, R. Searching for ten targets simultaneously. *Perceptual and Motor Skills,* 1963, **17**, 955–961.

NICKERSON, R. S. Response times with a memory-dependent decision task. *Journal of Experimental Psychology,* 1966, **72**, 761–769.

NORMAN, D. A. Acquisition and retention in short-term memory. *Journal of Experimental Psychology,* 1966, **72**, 369–381.

POSNER, M. I., & KEELE, S. W. Decay of visual information from a single letter. *Science,* 1967, **158**, 137–139.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs,* 1960, **74**(11, Whole No. 498).

STERNBERG, S. High-speed scanning in human memory. *Science,* 1966, **153**, 652–654.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

YNTEMA, D. B., & TRASK, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 65–74.

(Received February 17, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 238–241

LIST LENGTH AND SINGLE-TRIAL SHORT-TERM MEMORY[1]

DONALD A. SCHUMSKY,[2] ANTHONY F. GRASHA,[3] JOHN TRINDER
AND CHARLES L. RICHMAN[4]

University of Cincinnati

Each of 803 *S*s was required to learn a sequence of 3, 5, or 7 CVCs in one presentation. A group testing procedure was employed in which *S* recalled the first, middle, or last item in the list after either 0, 15, or 30 sec. Results revealed statistically significant effects associated with list length and serial position. No significant effects were associated with interval. The data were interpreted as suggesting that memory loss over interval does not appear unless *S* experiences a series of trials.

It is now a fairly well-established fact that the amount of memory loss in a short-term memory (STM) experiment can be a function of the ordinal position of the trial on which such loss is studied (Keppel & Underwood, 1962; Loess, 1964; Wickens, Born, & Allen, 1963). That is, retention when studied on the *n*th trial is not independent of previous tests and must be considered different from retention as studied on previous trials. Thus, the classic STM procedure of Peterson and Peterson (1959), which at the time seemed a minimal memory experiment, must be looked upon as a series of individual tests with changes occurring as a function of the ordinal position of those tests. It would follow then that some of the important findings in STM experiments can be obscured by a procedure which sums the data from many memory test trials.

Keppel (1965) and Schumsky, Grasha, Eimer, and Trinder (1967) argued that one test of a simple trace decay notion must involve a truly minimal memory experiment. This experiment must be a one-trial, single test of STM. In the Schumsky et al. experiment, *S*s were presented a single list of five items, one item at a time, and then recall was tested after 0, 15, or 30 sec. The *S*s were given one and only one trial and, at the end of the retention interval, were asked to recall either the first, third, or fifth syllable in the list. The results of that experiment revealed a significant effect associated only with serial position. No significant memory loss over the 30-sec. retention interval was obtained in that experiment. Schumsky et al. argued that their evidence offered no suggestion of trace decay. They suggested that for a trace decay notion to be consonant with the findings of their experiment, either trace decay must necessarily be a result of an overload in the STM storage system due to increased stimulus length, or the decay process must in some way be set in motion after some sort of proactive priming. The present experiment is an attempt to explore the effects of storage overload by varying list length in a single-trial–single-test experiment.

METHOD

Experimental design.—The *S*s were 803 undergraduate students at the University of Cincinnati. The basic design of the experiment was a 3 × 3 × 3 factorial varying list length (either 3, 5, or 7 Glaze high-M CVCs), retention interval (either 0, 15, or 30 sec.), and serial position of the item cued at recall (either the first, middle, or last item). Three replications of this basic experimental design were performed to allow for counterbalancing of specific test syllable over test position. Thus, by the end of the experiment, each syllable had been tested at each test position. The same test items were used in each list, with variation in the filler items providing variation in the length of list. The experiment employed a group testing procedure. Each group run represented one combination of list length and retention interval from each of the three replications. Use of a written cue (see

---

[1] This research was reported at the meeting of the Psychonomic Society, St. Louis, November 1968.

[2] Requests for reprints should be sent to Donald A. Schumsky, Department of Psychology, University of Cincinnati, Cincinnati, Ohio 45221.

[3] Now at the Staff and Faculty Building, United States Army Artillery and Missile School, Ft. Sill, Oklahoma.

[4] Now at Wake Forest University.

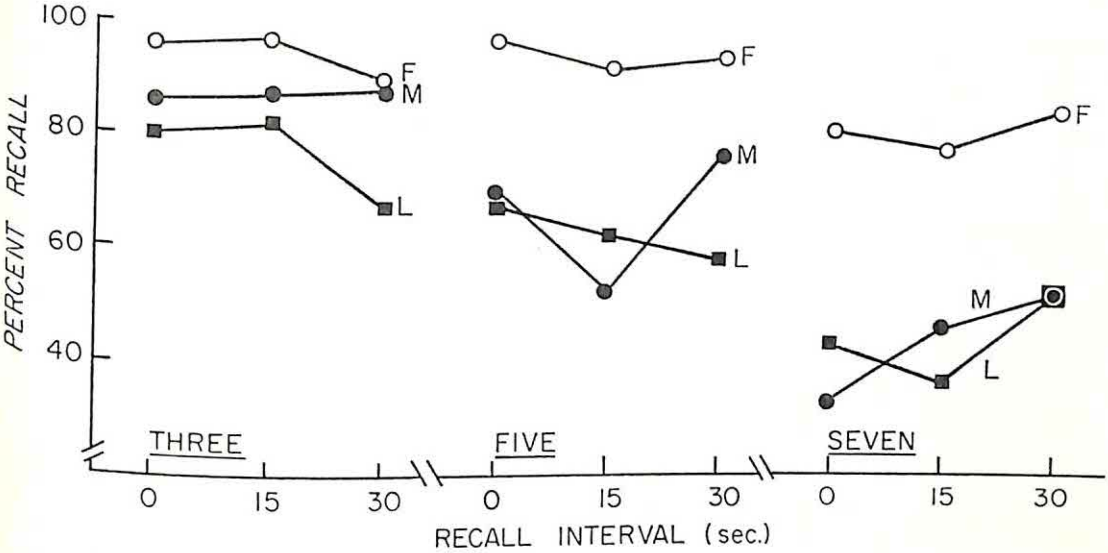


FIG. 1. Percentage of recall of first (F), middle (M), or last (L) stimulus for each of three list lengths as a function of retention interval.

the following section) allowed all serial positions to be tested at once.

Procedure.—After being seated in the experimental room, *S* was given a sheet of paper, face down, on which rows of random letters were printed. At the bottom of this page was a small rectangle in which *S* was to write his response. Stapled to the back of this sheet was another sheet of paper, folded so that *S* could not see what was written on it. This paper contained *S*'s cue for recall.

The *S*s were told that they were taking part in an experiment designed to study their ability to concentrate and that this ability would be assessed by determining how well they could check vowels from random strings of letters while being required to remember something else. They were told that a list of nonsense syllables would be presented one at a time. They were to memorize these items in order, and after a period of vowel checking, they would be asked to write all or part of the list.

Stimuli were presented visually, via an externally programmed Kodak Carousel projector, one at a time, each for 2 sec. A slide with the word "ready" appeared on the wall to start the trial. This was followed immediately by the list of either 3, 5, or 7 CVCs. Following the last CVC was a slide with 5 dots. This was *S*'s signal to turn over his sheet of paper and start checking vowels as "quickly and as accurately as possible." After the appropriate retention interval, *E* commanded *S* to "stop" checking vowels. For the 0-sec. retention interval *S*s, the command "stop" was given simultaneously with presentation of the dots. At this point, *S* turned over his sheet and unfolded the page with the cue which indicated that he was to write either the *first*, *middle*, or *last* syllable. He then turned back to the front of his sheet and wrote his recall.

RESULTS

The main findings of the present experiment are presented in Fig. 1 as the percentage of correct recall for each of the serial positions tested (F indicates recall of the first, M the middle, and L the last serial position) as a function of list length (3, 5, or 7) and retention interval (0, 15, or 30 sec.). Due to the group testing procedure, the points of Fig. 1 are based on different numbers of *S*s ranging from a low of 20 to a high of 42.

For purposes of statistical analysis, each of the 81 cells in the design (3 replications of the $3 \times 3 \times 3$ design) were treated as a single data point. This procedure solved the problem of unequal n in the cells and provided a more conservative analysis based on only 80 *df*. Thus, it must be noted that the universe of discourse, for purposes of inference, has been changed from *S*s to groups. The percentage of correct recall for each of the groups was subjected to an analysis of variance after arc sin transformation.

As can be seen from Fig. 1, there is little evidence of any retention loss over the 30-sec. retention interval for any combination of serial position and list length. This generalization is strongly supported by the

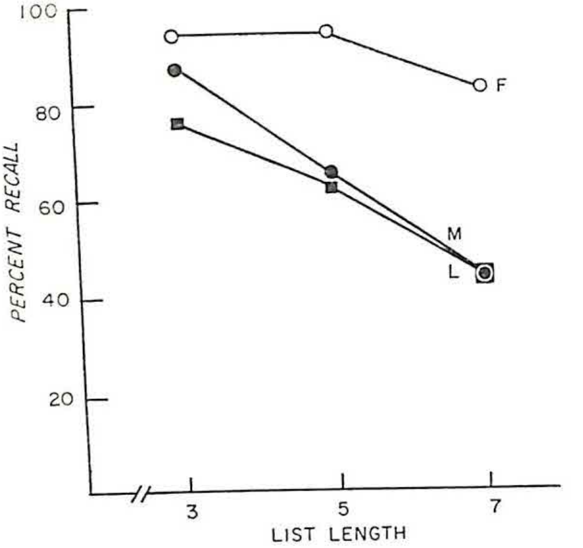


FIG. 2. Percentage of recall of first (F), middle (M), or last (L) serial position as a function of list length.

analysis of variance, where no mean square associated with the interval variable showed statistical significance (all F's $\leqq 1.00$). On the other hand, rather strong effects of both serial position and list length are evident, as can be seen from Fig. 1. There is an increasing degree of primacy as a function of list length. In addition, it is interesting to note that as list length increases, the performance gap between the first and middle stimulus in recall increases and the gap between the middle and the last stimulus seems to disappear. The effects of serial position, F $(2, 54) = 23.85$, $p < .001$, list length, F $(2, 54) = 18.98$, $p < .001$, and their interaction, F $(4, 54) = 2.65$, $p < .05$, were all significant.

Figure 2 presents the relationship between list length and recall summed over all retention intervals. It will be remembered that this was the only statistically significant interaction in the analysis of variance. Figure 2 provides rather strong demonstration of the increasing primacy with increases in list length. The effects of increased list length are most sharply seen in changes in recall of the middle position. The middle item when recalled as third item in a five-item list is not recalled as well as the last item of a three-item list. When recalled fourth in a seven-item list, the middle item is not recalled as well as the fifth item in a five-item list. In the seven-item list, the middle item is recalled no better than the last item despite the fact that its ordinal position is subspan (Position 4) vs. the superspan position of the last item (seventh). One possibility is that *S*s were confused at recall by failing to remember exactly which item was the middle one. If this were true, one would expect more intralist intrusions at the middle than at the end positions, particularly as list length increases. Error analysis revealed that by far the greater number of intrusions were intralist intrusions (over 90%). Although the percentage of intralist intrusions did increase with list length, this was not differential by serial position. Thus the error analysis fails to support a notion of confusion of position for the middle item. Unfortunately, these findings are also not consonant with a view explaining the reduction in recall of the middle position as a result of greater possibilities of bidirectional associational interference.

DISCUSSION

The results of the present experiment are in strong agreement with the results of the previous one-trial–one-test experiment of Schumsky et al. (1967). No loss of retention associated with retention interval was found in this or the previous experiment. Variables affecting recall are the traditional variables of list length and serial position. It would seem that even increasing list length to seven items, a value beyond the span for CVCs, resulted in no retention loss over retention interval. In a one-trial–one-test experiment of the sort reported here, retention interval per se leads to no significant memory loss and offers no support for a trace decay view of STM. This generalization seems to be consonant with the reports from other studies in the literature reporting one-trial data (Goggin, 1966; Keppel & Underwood, 1962; Rohrman & Jahnke, 1965).

In a previous study (Schumsky et al., 1967), some concern about the generality of the findings was indicated because the investigators were unable to find in the literature any other studies employing a similar recall procedure. Since the publication of that paper, there have appeared in the literature a few experiments employing a procedure involving the presentation of a list and cueing for recall out of that list by ordinal serial position (Murdock, 1968; Woodward & Murdock, 1968). Unfortunately, the data of the present experiment are not at all consonant with these reports. The strongest point of disagreement is the extreme recency effect noted in those experiments. In each of those experimental reports, the data presented are summed over many trials. The degree of recency found in such experiments may very well have been a function of the number of trials summed.

No support is offered for the finding, from studies of free recall, that an interpolated retention interval leads to differentially greater loss at the end than the beginning of a list (Glanzer & Cunitz, 1966; Postman & Phillips, 1965). This effect led Glanzer and Cunitz to attribute heightened recency effects in free recall to the action of an STM store. Glanzer and Cunitz pointed out that such recency effects should not be expected in studies employing ordered recall, as such recall forces a necessary delay and loss of items from an STM store. The failure to find the effect therefore may be due to an emphasis on order in the present experiment created by employing an ordinal cue for recall. On the other hand, as pointed out previously, other published studies employing an ordinal probe cue for recall do find pronounced recency effects. In addition, more recent research has indicated that delayed recall will produce recency losses in ordered recall experiments (Bartz, 1969; Jahnke, 1968). These studies summed data over multiple recall trials as did Glanzer and Cunitz. Postman and Phillips present data considering ordinal test position. These authors compare the differential retention effects on each list half over the testing sequence (first three tests vs. last three tests of six). Their data suggest an increase in the degree of recency loss as a function of ordinal position of test. It seems imminently possible that findings of sharp recency effects as well as the loss of these effects over a retention interval requires the summing of an extended sequence of trials. At the very least, *more than one trial seems necessary.*

In conclusion, the data of the present study offer no evidence of memory loss as a function of retention interval. In addition, no evidence for a Stimulus Length × Interval interaction is apparent. In reviewing the literature, it seems that evidence for "trace decay," loss of retention related to interval per se, requires an experiment employing a sequence of memory tests.

REFERENCES

BARTZ, W. H. Repetition and the memory stores. *Journal of Experimental Psychology*, 1969, **80**, 33–38.

GLANZER, M., & CUNITZ, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

GOGGIN, J. Retroactive and proactive inhibition in the short-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 526–535.

JAHNKE, J. C. Delayed recall and the serial-position effect of short-term memory. *Journal of Experimental Psychology*, 1968, **76**, 618–622.

KEPPEL, G. Problems of method in the study of short-term memory. *Psychological Bulletin*, 1965, **63**, 1–13.

KEPPEL, G., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

LOESS, H. Proactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 362–368.

MURDOCK, B. B., JR. Serial order effects in short-term memory. *Journal of Experimental Psychology*, 1968, **76**(4, Pt. 2).

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

POSTMAN, L., & PHILLIPS, L. W. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 132–138.

ROHRMAN, N. L., & JAHNKE, J. C. Effect of recall condition, presentation rate and retention interval on short-term memory. *Psychological Reports*, 1965, **16**, 877–883.

SCHUMSKY, D. A., GRASHA, A. F., EIMER, E. O., & TRINDER, J. Intra-list interference in single trial short-term memory. *Psychonomic Science*, 1967, **9**, 545–546.

WICKENS, D. D., BORN, D. G., & ALLEN, C. K. Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 440–445.

WOODWARD, A. E., & MURDOCK, B. B., JR. Positional and sequential probes in serial learning. *Canadian Journal of Psychology*, 1968, **22**, 131–138.

(Received February 17, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 242–249

DEVELOPMENTAL STUDY OF PERFORMANCE ON CONCEPTUAL PROBLEMS INVOLVING A RULE SHIFT[1]

DONALD E. GUY[2]

Wells College

In a conceptual rule learning task, four age levels (6, 7, 9 yr., and adult) attained a preshift solution, at which time each *S* was shifted to either a complemental (C) or noncomplemental (NC) rule. A C shift is analogous to a reversal (R) shift, requiring *S* to reverse all class assignments, while the NC shift is analogous to a nonreversal (NR) shift, requiring *S* to change only some of the stimulus category assignments. The relative performance levels on C and NC shifts paralleled those observed in earlier studies of R and NR shifts. Specifically, all *S*s except the 6-yr.-olds reached a higher level of performance on the C shift relative to the NC shift. The 6-yr.-olds found the C shift more difficult than the NC shift. The results show the comparability of developmental trends obtained in the two shift paradigms (C-NC and R-NR).

Any concept consists of two separable components: the relevant stimulus attributes and the rule specifying a relationship between the attributes. Most studies of conceptual behavior have dealt with attribute identification (AI), wherein the rule has constituted a given condition of the problem and *S* is required to discover the relevant attributes. There are, in contrast, only a few studies of the rule learning (RL) aspect of conceptual behavior, wherein the relevant attributes are known and the rule relating them must be discovered.

Evidence was reported by Haygood and Bourne (1965) that certain *S*s develop a stimulus coding strategy in the course of solving several RL problems. This strategy, which collapses the entire stimulus population according to whether relevant attributes are present (T) or absent (F), reduces the difficulty associated with learning and using various rules. Four resulting stimulus classes resemble the truth table of logic, e.g., stimulus patterns embodying both (TT), the first but not the second (TF), the second but not the first (FT), and neither (FF) relevant attribute. In RL problems, *S* learns to assign these four classes to two categories, positive or negative instances of the concept. Thus the strategy converts RL problems into an almost trivial 4:2 paired-associate task.

Performance in solution shift problems has been the object of study in many AI experiments. Reversal (R) and nonreversal (NR) are two basic types of solution shifts. The R shift requires *S* to reverse the attribute category assignments learned initially, while the NR shift introduces a different relevant dimension. For example, *S*'s task on the initial problem might be to sort red stimuli into Category A and green stimuli into Category B. At the R shift, response assignments would be changed to red into Category B and green into Category A. In contrast, the NR shift involves a change in dimensions, e.g., from the same initial solution to square into Category A and triangle into Category B. It is often reported that R shifts are more difficult than NR shifts for naive infrahuman *S*s and preverbal children, but that older children and adults tend to find the R shift easier. Even though adults solve both types of shifts more rapidly than preverbal children, an opposite ordering of shift difficulty for older and younger *S*s often obtains. Kendler and Kendler (1962) proposed that verbal mediational processes account for the better performance of older *S*s on R shifts.

---

[1] This research was undertaken in the Institute for the Study of Intellectual Behavior, University of Colorado, and is Publication No. 2 of the Institute. This report is based on a doctoral dissertation submitted to the University of Colorado. The author expresses his gratitude to Lyle E. Bourne, Jr., for his help in its preparation. This work was supported in part by National Science Foundation Grant GB-3404, by National Institute of Mental Health Grant MH 14314-01, by a United States Public Health Service predoctoral fellowship, and by the University of Colorado Council on Research and Creative Work. Appreciation is also extended to the Boulder Valley school district in Colorado and the Southern Cayuga school district in New York for their cooperation in providing *S*s for the study.

[2] Requests for reprints should be sent to Donald E. Guy, who is now at Central Washington State College, Department of Psychology, Ellensburg, Washington 98926.

The present experiment is the first to investigate performance on shift problems involving a change of rules rather than stimulus attributes. Two types of shifts were studied. In both, the relevant stimulus attributes were named for *S* at the outset and remained the same during both the preshift and postshift phases of the learning task. Some *S*s were shifted from one rule to its complement (i.e., mirror image), while others were shifted to a noncomplementary rule. The complemental (C) rule shift is analogous to the R shift in AI, for it involves a complete reversal of stimulus class to response category assignments. The noncomplemental (NC) shift is analogous to the NR shift because only some portion of the initial assignments are changed, i.e., only two of the four classes of the truth table are reassigned as either positive or negative instances.

Based on this analogy with AI tasks, it might be expected that C shifts will be easier than NC shifts for adults but more difficult for young children. To explore this possibility, C and NC shifts were administered to *S*s of four different ages in the present study.

Method

Subjects and design.—One hundred sixty *S*s participated in the experiment: 48 kindergarten children, 32 first graders, 32 third graders, and 48 college students 18-30 yr. of age. The mean ages for the three groups of children were 6.0, 7.0, and 9.0 yr., respectively. All of the children participating in the experiment were from the Boulder Valley school district, Colorado State, with the exception of 16 kindergarten children from the Southern Cayuga school district in New York State. The adults were students from the University of Colorado, with the exception of 16 students from Wells College. The 32 additional *S*s from New York (16 kindergarten and 16 adult) were run as a reliability check to the original data. The experimental design was a $2 \times 4 \times 2 \times 4$ factorial consisting of two different problems, i.e., different pairs of relevant attributes, four ages, two types of rule shifts, i.e., C and NC, and four different rules. The rules selected were conjunction and inclusive disjunction and their respective complements, alternative denial and joint denial. The manner in which these rules assign stimulus classes to response categories is shown in Table 1. The eight rule shift conditions used in the study are given in Table 2.

Materials and apparatus.—The stimuli consisted of geometric patterns, prepared on 3×5 in. cards, which varied along two dimensions with three levels per dimension. Dimensions and their values were: form (square, triangle, and circle) and color (red, yellow, and blue). Forty cards were arranged in a predetermined order allowing for the basic deck of nine paterns to be repeated several times. This procedure served to eliminate interruptions and insured that the four stimulus contingencies (TT, TF, FT, and FF) were equally represented in all tasks. The *S* was seated at a table facing *E*. A divider was used to prevent *S* from viewing the decks of cards needed for the

TABLE 1

Assignments of Stimulus Classes to Response Categories (+ and −) under the Four Rules Used in the Experiment

Stimulus class	General notation	Stimulus set	Rule			
			Conjunction	Alternative denial	Inclusive disjunction	Joint denial
RS	TT	RS	+	−	+	−
$R\bar{S}$	TF	RTr, RC	−	+	+	−
$\bar{R}S$	FT	YS, BS	−	+	+	−
$\bar{R}\bar{S}$	FF	YTr, YC BTr, BC	−	+	−	+

Note.—T = true (or present); F = false (or absent); R = red; Y = yellow; B = blue; S = square; Tr = triangle; C = circle.

TABLE 2

Eight Rule Shift Conditions

Preshift problem	Postshift problem	Shift type
Conjunction	Alternative denial	C
	Joint denial	NC
Disjunction	Joint denial	C
	Alternative denial	NC
Alternative denial	Conjunction	C
	Disjunction	NC
Joint denial	Disjunction	C
	Conjunction	NC

rule shift. Upon solution of the preshift problem, *E* would begin a new deck for the C or NC shift. A total of five decks (40 cards in each) were arranged for *E* to present. Two small cardboard boxes with slots were used as Categories A and B. Presentation of the stimulus patterns and recording of responses were performed manually.

Task and procedure.—Each *S* was required to sort a series of visually presented stimuli into two categories, the positive and the negative instances of a concept. The correct classification in any problem was determined both by a pair of relevant attributes (one from each of the two dimensions) and by a conceptual rule specifying the relationship between the attributes. Two such problems (with the same relevant attributes but different rules) were administered without interruption.

Pretraining.—All *S*s were pretrained in the "truth table" strategy. Detailed oral instructions described the stimulus population, and each *S* was required to verbalize whether the relevant (given) attributes were present or absent in each pattern that was categorized. The 6-, 7-, and 9-yr.-old groups were told that they were going to play a game. They were to pretend that the cards had a first and a last name, color being the first and form the last name. Thus, a card might be named "yellow circle." A small open box with four panels (2 × 2) was used to allow *S*s to sort 20 patterns into the proper stimulus contingencies (TT, TF, FT, and FF). The pretraining for the 6-yr.-olds consisted of verbalization training only, and the procedure was modified in an attempt to reduce the difficulty of the task for the six-yr.-old *S*s. In particular, only two levels of the two dimensions were used (red and yellow for color, and circle and squares for form). Moreover, the same attributes used in *S*'s pretraining task were used for his rule transfer problem. Each child practiced "naming," i.e., whether both attributes were present or absent for 20 patterns, prior to the experiment proper. Each panel was labeled according to whether the first or second attribute was present or absent. For example, if *S* was required to sort stimuli according to the relevant attributes "yellow circle" and was confronted with a yellow square, he would say, "yellow, not circle," and place it in the TF category marked "yellow only." One pretraining problem was used.

The next phase of the pretraining for all but the 6-yr.-old *S*s demonstrated how the four contingencies could be collapsed to two categories, positive or negative instances. Two unlabeled boxes (categories) were displayed and a two-category sorting procedure was illustrated using both conditional and biconditional rules (see Haygood & Bourne, 1965) in order to avoid direct practice on rules used in the experiment proper. The conditional rule is verbalized "if X then Y," which implies that a stimulus that is X but not Y is not an example of the concept; all others are. In other words the contingencies TT, FT, and FF are positive, and TF is negative. The biconditional rule is stated "X if and only if Y," which means that a stimulus must have both X and Y or neither X nor Y in order to be an example of the concept. The contingencies TT and FF are positive, while TF and FT are negative. Without naming the rule, *E* showed *S* how the four classes he had just sorted could be placed into Category A and Category B. The *S* then sorted 10 cards according to the conditional rule and the remaining 10 cards according to the biconditional rule with *E* prompting *S* as necessary. When *E* was certain that *S* understood the truth table collapsing scheme, the main phase of the experiment began.

Rule transfer.—The *S*s were told that one stimulus pattern at a time would be shown to them and that their task was to place it correctly in one of the two boxes. The relevant attributes were listed for *S* (either "red, square" for Problem A or "blue, triangle" for Problem B) and left exposed throughout the experiment. The kindergarten children had "yellow, circle" for Problem B. The *S*s were further instructed that they would be told right or wrong by *E* for each response and that for every correct response they would receive a marble. Children were told that they could trade the marbles in for one, two, or three toys depending on how well they did. Each toy was worth from 30¢ to 50¢. The adults were told that the study was concerned with using marbles as feedback, but they could win $1, $2, or $3 depending on their performance. Each *S* was assigned to one of the eight rule transfer conditions. When the criterion of solution for the problem was reached (12 consecutively correct trials), the solution was shifted to a new rule, either a complement or a noncomplement. Performance continued until 12 patterns were correctly identified in succession.

Results

Analyses of variance performed on errors and trials to solution of the preshift problem indicated that only the age effect was significant, *F* (3, 144) = 7.3 for trials and 5.38

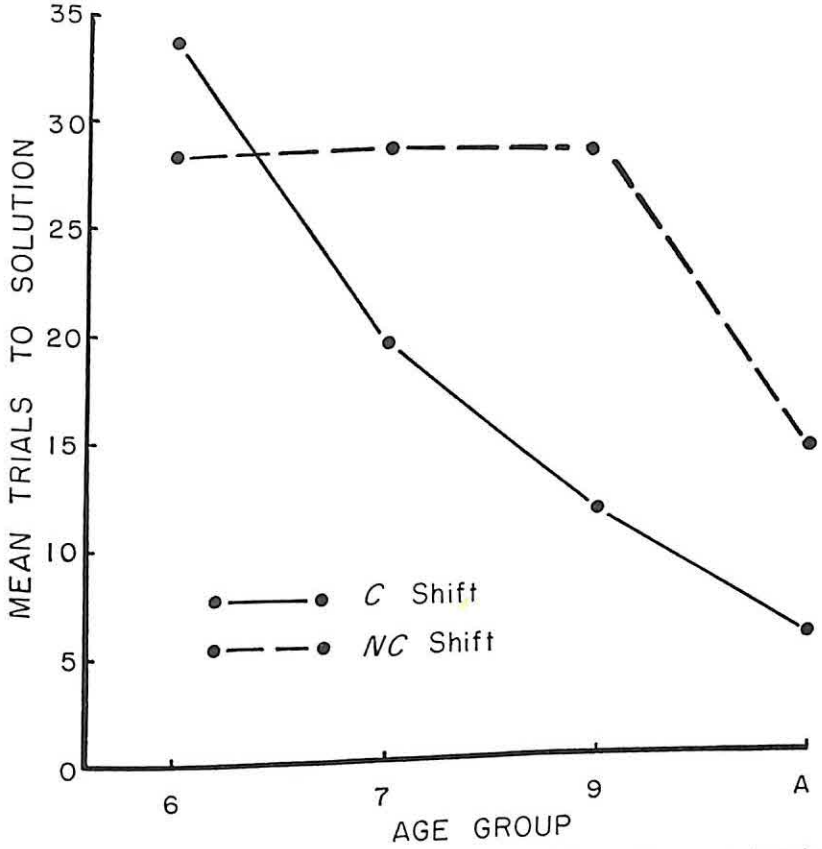


FIG. 1. Mean number of trials to solution in the postshift problem as a function of age and the type of shift required in the postshift.

for errors, $p < .01$. Mean trials to criterion for the preshift problem were 28.0, 25.8, 20.6, and 9.4 for ages of 6, 7, 9 yr., and adult, respectively. This result supports the notion of increasing levels of conceptual skill across the ages from 6 yr. to adult and implies also that groups, otherwise treated identically, performed at essentially the same level.

Mean trials to solution on the postshift problem are plotted as a function of age and the type of rule shift in Fig. 1. The C shift was easier than the NC shift for all but the youngest age group. Analyses of variance on these data confirmed the reliability of this difference. The shift variable (C vs. NC) provided significant *F*s for trials and errors, $F(1, 152) = 9.22$ and 4.02, $p < .01$ and .05, respectively. The age effect was reliable, $F(3, 152) = 12.65$ (trials) and 10.35 (errors), $p < .01$. The Age × Type of Shift interaction also yielded significant *F*s for trials and errors, $F(3, 152) = 3.60$ and 4.67, p's $< .05$, respectively. There were no reliable differences either between problems (A or B), rules at each of the four age levels, or sets of complemental rules (conjunctive and alternative denial vs. disjunctive and joint denial). The *S*s from the two geographical locations yielded comparable data, as evidenced by the similarity in mean trials to solution (11.3 and 8.0 for adults and 30.6 and 31.5 for 6-yr.-olds from Colorado and New York, respectively) for both C and NC shifts averaged together.

Kendler and Kendler (1962) have reported that those *S*s who reach solution more rapidly in an initial problem tend to solve R shifts more easily than NR shifts. In order to determine whether *S*s in the present experiment differed on the postshift problem as a function of preshift performance, *S*s were divided into subgroups at the median error score on the initial problem

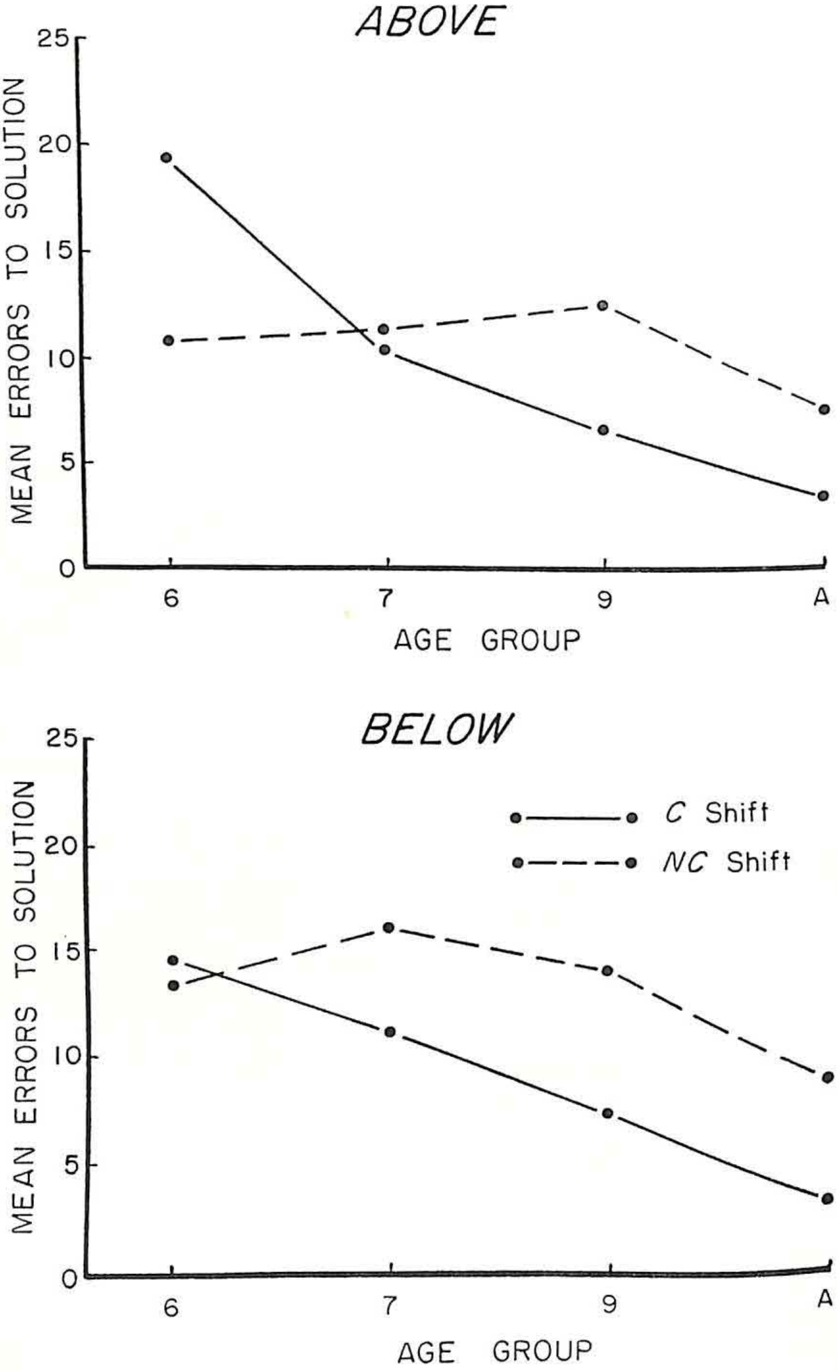


FIG. 2. Mean errors to solution in the postshift problem for *S*s above and below the preshift error median.

within each shift condition (C vs. NC). Figure 2 illustrates that for the "better" *S*s (those below the preshift error median), a C shift is noticeably easier than the NC shift for all but the 6-yr.-olds, who demonstrated almost equal difficulty with both shifts. On the other hand, the "poorer" 6-yr.-olds appeared to have much more difficulty with the C type of shift, giving mean error scores of 19.25 and 10.83 for the C and NC types, respectively. The analysis of variance performed on postshift errors to solution for

*S*s above and below the preshift median showed that age and type-of-shift effects were significant at the .05 and .01 levels, respectively. The above–below median effect was unreliable, but the Age × Above–Below Median interaction was significant, $F\ (1, 84) = 5.55$, $p < .05$. The Type of Shift × Age interaction as well as the third-order Above–Below Median × Type of Shift × Age interaction were reliable, $F\ (3, 144) = 41.20$ and 30.43, p's $< .01$, respectively. The Above–Below Median × Age interaction was unreliable. Six-yr.-old *S*s performing above the preshift median did reliably better on the NC shift than the C shift, $t\ (22) = 2.91$, $p < .01$. There was also a significant difference on the C shift between means for 6-yr.-old *S*s above and below the preshift

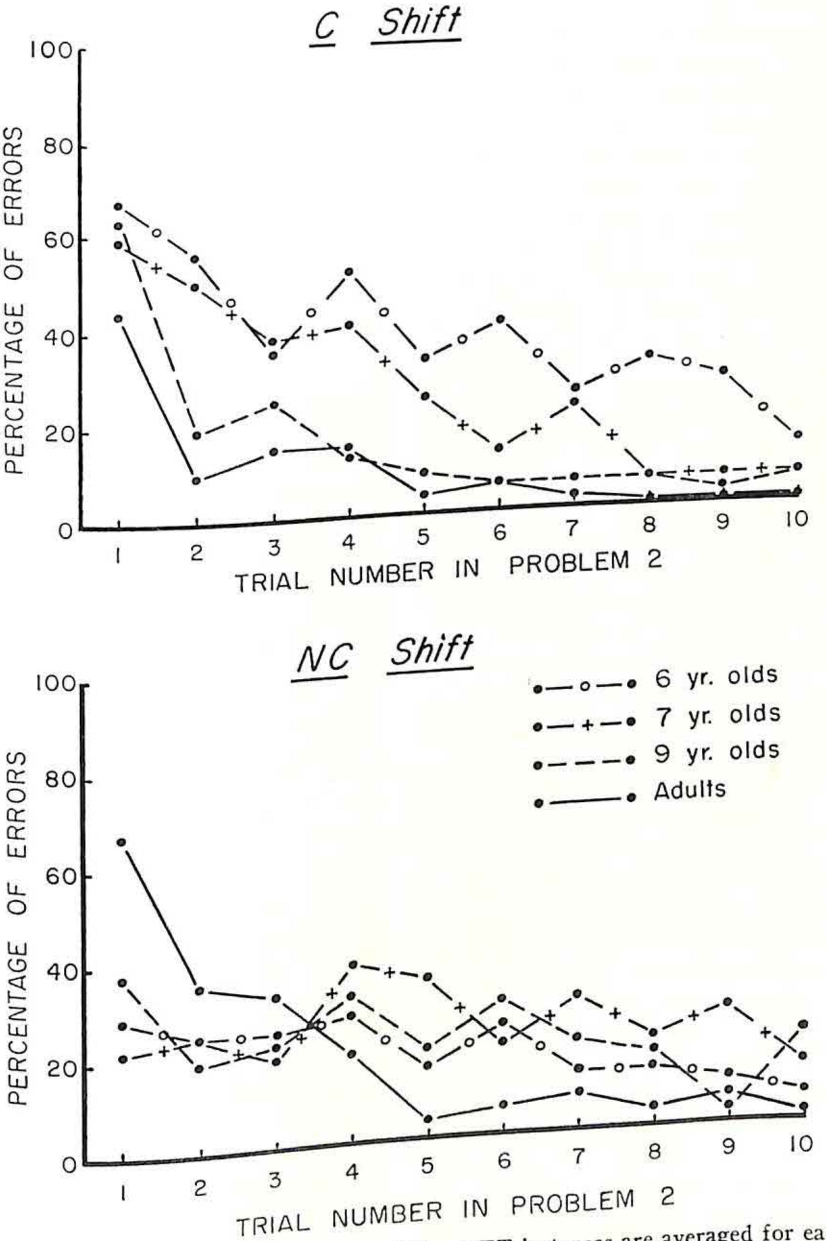


Fig. 3. Learning curve plotted forward, where TF and FT instances are averaged for each age group. (Each plotted point indicates the total number of errors divided by the appropriate number of *S*s for each condition.)

median, $t\ (22) = 2.23$, $p < .05$. These results demonstrate a C-NC inversion of difficulty for younger *S*s of poorer conceptual ability.

To determine whether learning the assignment of certain stimulus contingencies to response categories occurred more rapidly than others, computations on postshift data were performed on (*a*) the mean number of examples of each class presented prior to *S*'s last error *on that class,* (*b*) the mean number of errors for each class, and (*c*) the probability of an error on each class (the ratio of *b* to *a*). Perhaps the most striking feature of these analyses was the comparison of the four contingencies between the two types of shifts. Twice as many errors were made on all classes for the NC shift relative to the C shift for all but the 6-yr.-olds. However, the probability of an error on any class (the average of all four classes) was nearly the same, .54 and .49 for C and NC shifts, respectively.

The fact that 26 out of 48 adult *S*s solved the postshift problem in the minimal number of trials (number of trials to present one example of each class) suggests that they learned to use the truth table principle more quickly than children. In contrast, only 3 out of 48 6-yr.-olds, 3 out of 32 7-yr.-olds, and 5 out of 32 9-yr.-olds solved the postshift problem in the minimum number of trials. If *S* has some natural tendency to reverse, C-shift *S*s might make only one error in the postshift problem. The NC *S*s, on the other hand, might be expected to need to see one example of each of the four contingencies. In the C-shift conditions, 10 out of 24 adults solved the postshift problem after making one error. Of the 10 *S*s who solved the NC postshift problem with four errors or less, 6 were also adults. Of the 26 adults solving the postshift problem with four errors or less, 20 had the C shift. This evidence might be taken to suggest that adults have both better developed conceptual skills and a stronger tendency to reverse than do children.

Learning curves for the various classes of stimuli were plotted separately for the two types of shifts by adding errors on the TF and FT classes and on the TT and FF classes. These pairs of classes were combined in constructing curves because the TF and FT classes in the NC shift and TT and FF classes in both shifts were assigned to the same categories. A crucial difference between adults and children is illustrated in Fig. 3 for the NC shift on the TF and FT classes. For all *S*s the category assignment of these instances remained the same in both problems. (TT, FF, FT, TF was the sequence of stimuli on the first four postshift trials for all *S*s.) The first trial on which *S* encountered an unchanged instance was Trial 3 (FT) after the NC shift. From Fig. 3 it is clear that adults took the first two trials, which infirm their Problem 1 solutions, as a cue to reverse all categories since 67% of the adult *S*s made an error on Trial 3; only 29% of the children made an error on this trial. A *Z* test comparing these two percentages was significant, $Z = 4.50$, $p < .001$. After the initial error on the FT, however, the reversal tendency was reduced, and by the fifth time either an FT or TF class appeared, only 4% of the adults made an error. Children made 20–30% errors the first four times these classes appeared, and their error rate dropped only gradually thereafter.

Both C and NC shift learning curves for TT and FF instances showed a marked discrepancy between children and adults on the early performance trials. Adults obtained around 54% errors on the second trial of both shifts, while children obtained around 84% errors. A rapid drop in errors occurred for the adults (to 6%) by the fourth presentation of the TT and FF class, while the 6- and 7-yr.-olds dropped only to 40% and reduced their subsequent percentage of errors on these classes relatively slowly. Nine-yr.-olds, on the other hand, began with 80% errors on both shifts and dropped to 13% on the third trial of the C shift, but only to 50% on the third trial of the NC shift.

Evidence from the remainder of the learning curves supports the notion that on a shift problem, adults attempt more quickly to reverse the response categories than children (instead of attempting new associations for all four classes).

DISCUSSION

The results of the experiment indicate developmental trends similar to those obtained in attribute learning studies. The relative difficulty of C and NC shifts changes with age from NC easier than C for 6-yr.-olds to C easier than NC for adults. Even 7-yr.-olds, however, perform better on the C shift. One result that helps to clarify that outcome is shown in the analysis of "better" and "poorer" *S*s (see Fig. 2). While the C shift was easier than NC for 9-yr.-olds and adults, regardless of the level of their overall performance, the difficulty of these shifts was about the same for the poorer 7-yr.-olds and clearly reversed for the poorer 6-yr.-olds. Only the better 7-yr.-olds found the C shift easier than the NC. The better 6-yr.-olds, on the other hand, displayed equal difficulty with both shifts.

It is unusual to find no significant differences in difficulty among rules. This finding conflicts with earlier research (e.g., Haygood & Bourne, 1965; King, 1966). One possible explanation for the equality of rules lies in the truth table pretraining procedure used in the present experiment. A study by Haygood and Kiehlbach (1965) demonstrated a reduction in the difficulty of all rules as a result of truth table experience. Maximally efficient performance was achieved by all *S*s, adults in this case, on both the conjunctive or disjunctive rule after truth table pretraining. Because pretraining was used in all conditions of the present experiment, the results suggest that its effect in equalizing rule difficulty probably holds for children as well as adults. The verbalization training on the truth table for the 6-yr.-olds may have been sufficient to eliminate rule differences that may have otherwise obtained.

A study by King (1966) showed that the absolute difference in difficulty of conjunctive and disjunctive is larger for 7-yr.-olds than for older *S*s. However, by the third problem in his study, rule differences were negligible. This result held for all age levels in the present experiment. Apparently, practice over three problems with the same rule in King's experiment had the same equalizing effect for rules that training with the truth table did in the present case.

The results show that maximally efficient performance was more common among the older *S*s and on C-shift problems. There appeared to be a marked tendency for adults to reverse the category responses immediately after the first infirming error trial. Only three 6- and 7-yr.-olds and five 9-yr.-olds approached this level of performance. The reversal tendency for adults is also represented in Fig. 3. Adults have a .67 probability of making an error on the first presentation of a TF or FT instance, while 6-, 7-, and 9-yr.olds have only a .29, .22, and .38 error probability, respectively. Adults reverse the TF and FT classes even when the assignment in the postshift problem is the same as in the preshift problem. The difference between probabilities is significant and demonstrates that adults only need one infirming trial subsequent to a series of correct responses before they reverse their category responses. The suggested conclusion is that all class assignments do not have to be demonstrated to *S* before he attempts the reversal. Contrasted to this, the children are more likely to test each class for confirmation before any change in category assignment is made. The tendency to reverse at the first error on the postshift problem is probably a function of both age and experience and is related in an obvious way to solution shift behavior in AI tasks. As such, the data are clearly consistent with parallel studies within the AI paradigm.

REFERENCES

HAYGOOD, R. C., & BOURNE, L. E., JR. Attribute and rule learning aspects of conceptual behavior. *Psychological Review*, 1965, 72, 175–195.

HAYGOOD, R. C., & KIEHLBACH, J. B. Effects of logical pretraining on concept rule-learning performance. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, April/May 1965.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, 69, 1–16.

KING, W. A developmental study of rule learning. *Journal of Experimental Child Psychology*, 1966, 4, 217–231.

(Received February 17, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 250–257

ANALYSIS OF SEQUENTIAL EFFECTS IN CHOICE REACTION TIMES[1]

ROBERT J. REMINGTON[2]

Michigan State University

A comprehensive method of sequential data analysis is introduced and applied to reaction time (RT) data from a choice reaction experiment. The choice reaction time (CRT) experiment presented in the present paper consisted of two experimental conditions: (*a*) a two-choice condition in which the stimulus events were equiprobable and (*b*) a two-choice condition in which one of the stimulus events appeared with a probability of .70. Each of five *S*s performed under both conditions. The results of the present study, in addition to demonstrating the feasibility of the new detailed approach to the study of sequential effects, revealed new information regarding the nature of the "repetition effect," i.e., the observed faster RT to repeated signals than for changed signals. It was found that the repetition effect could not be associated with a single repetition of a signal as previously supposed. Findings are reported that indicate how higher order sequential effects operate to produce the second-order repetition effect. The findings indicate that an adequate model of CRT must account for at least fifth-order sequential effects.

Most of the research on sequential effects in reaction time (RT) data centers around the phenomenon known as the "repetition effect," i.e., the often reported finding that the RT to a repeated signal is faster than the RT for a changed signal. Since the repetition effect was first reported by Hyman (1953), there have been a number of studies designed to provide some understanding of the nature of this effect (Bertelson, 1961, 1963, 1965; Hale, 1967; Hannes, 1968; Kornblum, 1967; Smith, 1968). These studies have shown that factors such as signal probability, S-R compatibility, intertrial interval, number of alternatives, response mode, and S-R mapping influence the magnitude of the repetition effect differentially. Only a few researchers have gone beyond the simplest sort of sequential analysis associated with the repetition effect to explore higher order sequential effects in choice reaction time (CRT) data (Bertelson, 1961; Falmagne, 1965; Leonard, Newman, & Carpenter, 1966). The results reported by these researchers provide good evidence that higher order sequential effects do exist and that sequential effects in RT data cannot be adequately described by a consideration of interaction between adjacent trials only. However, only effects associated with a relatively small number of the possible higher order stimulus patterns were subjected to analysis by the latter researchers (i.e., repetition stimulus sequences and sequences in which successive presentations of a given stimulus are separated by intervals of varying size).

It is the major premise of the present paper that the data from a CRT experiment contain more information concerning the microstructure of underlying choice reaction processes than can be obtained through previously reported sequential data analysis procedures. In keeping with this premise, a comprehensive method of sequential data analysis was introduced and applied to RT data collected from a binary choice reaction experiment.

Discussion of the data analysis procedure employed in the present paper is facilitated by the introduction of some symbols and terminology. Let A represent the stimulus event appearing on Trial *n* (i.e., a first-order stimulus pattern). By looking at the stimulus events that can occur on Trial *n* and Trial $n-1$, two *second-order* stimulus patterns emerge: (*a*) the BA pattern (i.e., the stimulus event appearing on Trial $n-1$ is different from the stimulus event appearing on Trial *n*) and (*b*) the AA pattern (i.e., the stimulus event appearing on Trial $n-1$ is identical to the stimulus event appearing on Trial *n*). The simplest sort of sequential analysis of CRT data consists of comparing the mean RT for all stimulus events that are immediately preceded by the same stimulus event (the AA pattern) with the mean RT for all the stimulus events that are immediately preceded by a different stimulus event (the BA pattern). This particular comparison represents the extent of the sequential analysis carried out by a majority of the researchers who have reported on sequential effects in CRT data.

---

[1] This paper is based on a PhD dissertation submitted to Michigan State University. The author is greatly indebted to his major professor, Terrence M. Allen, for his interest, encouragement, and helpful suggestions.

[2] Requests for reprints should be sent to Robert J. Remington, who is now at International Business Machines Corporation, P.O. Box 12195, Research Triangle Park, North Carolina 27709.

In carrying the sequential analysis a step further, one would first delineate the *third-order* stimulus patterns. Examination of the stimulus events associated with Trials *n*, $n-1$, and $n-2$ reveals four possible third-order patterns: AAA, BAA, BBA, and ABA. The tree diagram in Fig. 1 illustrates how the various stimulus patterns are delineated as one considers successively higher orders through the fifth order.

It is now possible to associate a mean RT with each of the 31 stimulus patterns presented in Fig. 1. The overall mean of an RT distribution resulting from an *n*-trial CRT experiment would be the value associated with the first-order stimulus pat-

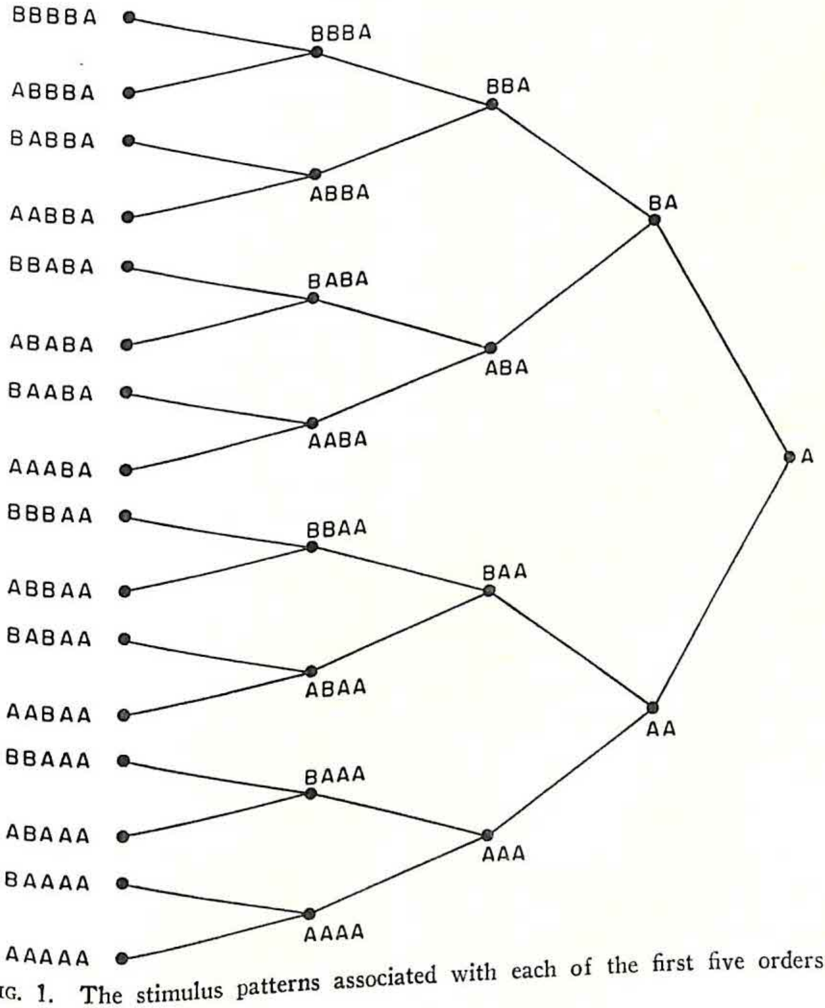


FIG. 1. The stimulus patterns associated with each of the first five orders.

tern A. Let this value be denoted as RT(A). Then, RT(AA) would represent the mean RT for all the stimulus events in the *n*-trial experiment that are immediately preceded by the same stimulus event (the mean RT for repetitions), and RT(BA) would represent the mean RT for all the stimulus events in the experiment that are immediately preceded by a different stimulus event (the mean RT for nonrepetitions). In short, the proposed tree diagram analysis of CRT data consists of finding the mean RT for each of the 31 stimulus patterns shown in Fig. 1.

The fragmentary nature of the approaches taken by the various investigators in their study of sequential effects in CRT performance becomes very apparent when the amount of information that they extract from the data is related to the potential information made available through the tree diagram approach. For example, the position in run analysis of repetitions introduced by Bertelson (1961) delineates only part of the lower branch of the tree diagram structure, while Falmagne's (1965) repetition analysis delineates only 4 scattered nodes within the tree diagram structure. In fact, only 12 of the 31 nodes of the tree diagram shown in Fig. 1 would be determined if all the conventional data analysis procedures were applied to a set of CRT data. How important are each of the remaining 19 stimulus patterns in determining RT? It seems reasonable that information concerning the influence of these unexplored stimulus patterns on RT would be useful in testing present theoretical models. It would also appear that the proposed tree diagram analysis, which incorporates all the sequential information (through the fifth-order) under one experimental design, represents a relatively comprehensive approach to studying the microstructure of choice reaction processes.

METHOD

Subjects.—Five volunteer *S*s (three females and two males) participated in the experiment. The *S*s ranged in age from 15 to 28 yr. They were previously unacquainted with task and had no knowledge of the aim of the experiment.

Apparatus.—The signals were presented by means of a visual display. The signals themselves consisted of four lights (No. 47 bulbs behind frosted lenses) approximately 1 in. in diameter mounted in a small metal cabinet positioned about 4 ft. in front of *S*. The stimulus display also contained a red warning light and a small green light used to provide *S* with knowledge of a correct response.

The response device was a small keyboard consisting of four typewriter keys. The keys were numbered, starting on the left, 1–4, corresponding spatially with the numbered stimulus lights. RT was measured from the onset of a stimulus light to the depression of one or more of the response keys. The depression of a key following a stimulus light was followed by the automatic recording of the RT with millisecond accuracy, stimulus number, and the specific key(s) depressed by *S*. A fixed 1-sec. delay between the onset of the warning light and the onset of the stimulus light (foreperiod) was employed throughout the experiment.

The control and recording portions of the equipment, located in a room adjacent to the room containing the visual display, allowed for automatic data collection without the necessity of visual inspection or intervening operations by *E*. Stimulus presentations were automatically controlled by means of punched paper tape programs read by a paper tape reader. The relevant information for each trial (i.e., a record of the stimulus light presented, the response made by *S*, and the RT in milliseconds for the trial) was automatically recorded in digital form onto paper tape by means of an eight-channel paper tape punch. The entire interstimulus interval, consisting of (*a*) the foreperiod, (*b*) the stepping of the control tape, (*c*) the CRT, and (*d*) the data recording, was about 4 sec.

Stimulus sequences.—All of the stimulus sequences employed in the present experiment were 200 trials long. Randomization of the stimulus events was achieved by means of a computer program. Additional structural characteristics of these stimulus sequences are presented in the following description of the experimental conditions:

1. *50:50 condition.* Four different binary sequences were constructed for this experimental condition. Under the 50:50 condition, both stimulus alternatives, consisting of Light 2 and Light 3, were presented equally often in each of the four 200-trial stimulus sequences.

2. *70:30 condition.* The second condition also employed a binary-choice reaction task. For this condition, four different stimulus sequences were constructed so that the proportions of alternatives were 70:30. In two of these 200-trial series, Stimulus Light 2 appeared 140 times, while in the other two series, Stimulus Light 3 appeared 140 times. Second- and third-order estimates of

redundancy for the combined stimulus sequences were nearly zero for both two-choice conditions.

Procedure.—Each *S* performed under both experimental conditions. At the beginning of the first experimental session, each *S* was familiarized with the general nature of the choice reaction task and the apparatus. During a typical experimental session, *S* was presented with four 200-trial stimulus sequences. Before each sequence, *S* was told the probability associated with the appearance of each of the possible stimulus events in terms of percentages. The *S*s were instructed to respond as quickly as possible to the presentation of the stimulus light without making more than 5% error responses during a given series. During all testing, *S*s wore lightweight dual headphones with ear-encompassing cushions that blocked out most of the external noises. After each sequence, *S* was given a brief rest period and instructions concerning the upcoming sequence. Under both experimental conditions the index finger on the left hand was used to respond to Light 2 and the index finger on the right hand to respond to Light 3.

General data analysis.—A total of 12,000 data records were collected during the course of the present experiment. Since early trials on CRT tasks typically display a relatively high degree of variability, which tends to obscure possible effects, a decision regarding which portion of the early data will be excluded from analysis is generally required. In the present study a separate decision was made for each experimental condition on the basis of sequential analyses made on the RTs from each successive 200-trial sequence or block for a given condition. That is, a separate detailed sequential analysis was made on the 1,000 RTs collected for all *S*s combined (200 RTs per *S*) on their first 200-trial block under a particular condition; a separate detailed sequential analysis was made on the 1,000 RTs collected for their second 200-trial block under the same condition; etc.

A relatively stable pattern of higher order sequential effects was observed for the last four blocks for the 50:50 condition. The first two blocks for this condition, being the first two blocks of practice on a CRT task for all *S*s, were characterized by rather erratic higher-order sequential effects as well as a large drop in overall mean RT. Therefore the first two blocks (i.e., the first 400 CRTs for each *S* under the condition) were excluded from the final data analysis for this experimental condition. Only the first block of trials for the 70:30 condition displayed an unstable pattern of higher order sequential effects. With the first block excluded from analysis, 5,000 RTs from the last five blocks remained to be analyzed. The RTs associated with error responses (approximately 1% of the total number of responses) were included in the final data analysis. Therefore the analyzable RT data consisted of 800 data records for each of five *S*s under the 50:50 condition and 1,000 data records from each *S* under the 70:30 condition.

The desired detailed sequential analysis of these data was only feasible with the aid of an electronic computer. The data, being punched onto paper tape in digital form, were read directly by the paper tape reader associated with the computer system located at the Michigan State University computer center. A computer program was developed which is capable of calculating and printing a number of statistics (mean RT, variance, standard deviation, number of observations, standard error of the mean, skew, and percentage of error responses) associated with every possible tree diagram node through a given order for (*a*) an overall tree diagram analysis and (*b*) a tree diagram analysis for each signal separately.

RESULTS AND DISCUSSION

The results of the tree diagram analysis of the 50:50-condition data are presented in Fig. 2. As indicated previously, this analysis associated a mean RT with every possible stimulus pattern through a given order (the fifth order in this case). Perhaps the most striking result was the overall picture that emerged when the tree diagram analysis was applied to the data. The graph in Fig. 2 clearly conveys the fact that stimulus patterns of a higher order than the popular second order have a reliable effect on RT. The results of this tree diagram analysis, represented in graphic form, also shed light on the nature of the repetition effect and some of the theoretical misconceptions associated with this popular second-order sequential effect.

It is generally believed that the repetition effect is the result of some sort of perceptual or response facilitation brought about by having seen or responded to the same stimulus event on the *immediately preceding* trial. A very simple test of this theoretical notion can be made by examining the mean RTs associated with the third-order components of the AA pattern (i.e., BAA and AAA). From the results reported in Fig. 2, it can be seen that the decrease in RT associated with the AA pattern (the repetition effect) can be attributed mainly to the AAA pattern and its higher order components and that RT(BAA) is nearly the same as the overall mean RT for the stimulus series. The important finding is that, at least for the 50:50 condition, the repetition

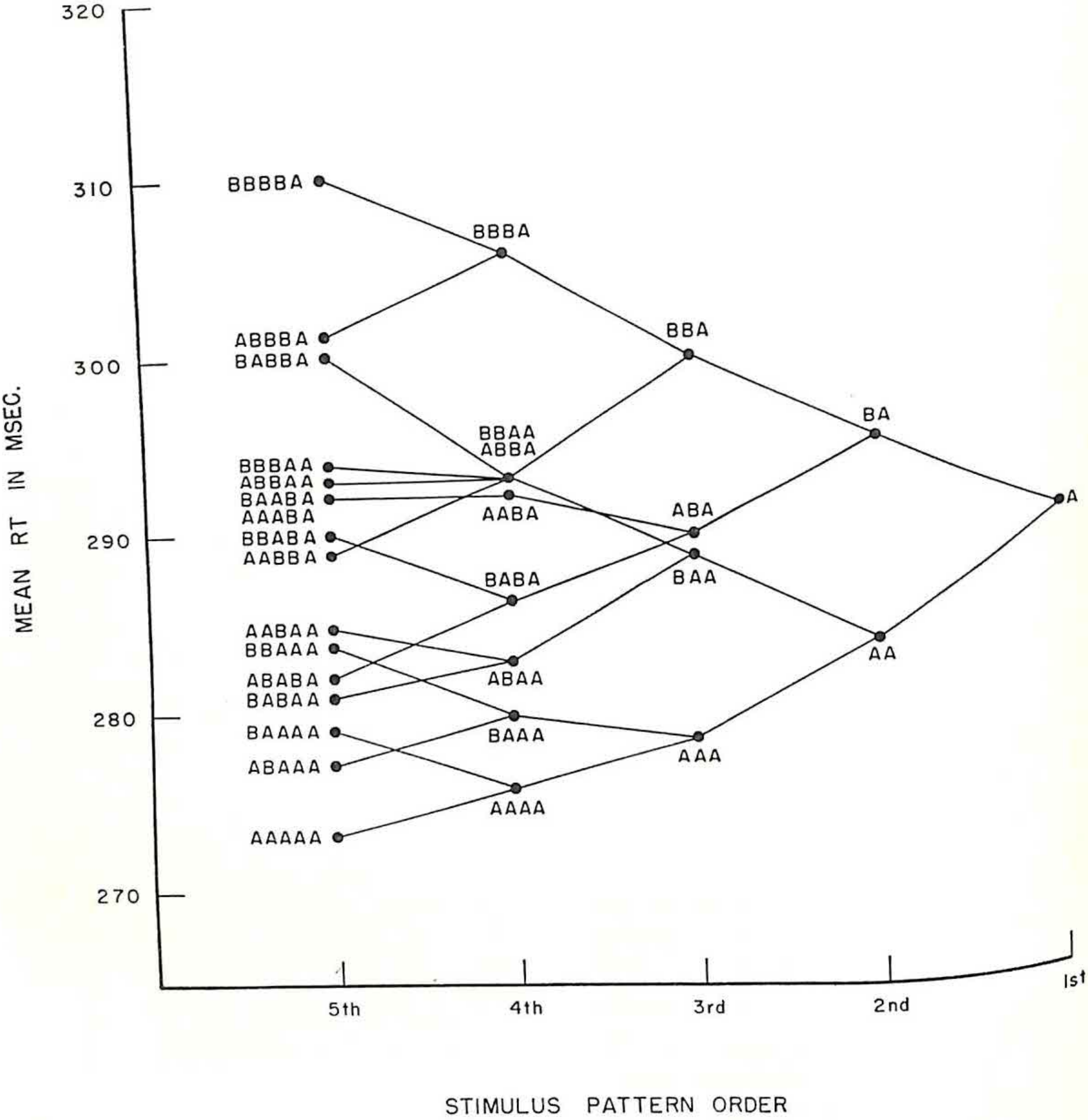


Fig. 2. Tree diagram analysis for the 50:50 condition.

effect could not be associated with a single repetition of a stimulus event.

Examination of the mean latencies associated with the third-order components of the BA pattern revealed more information concerning the nature of the repetition effect. Assuming that on Trial n, S is operating only on the information processed on Trial $n - 1$, there would be no reason to expect a difference between the mean latency associated with the ABA stimulus pattern and the mean latency associated with the BBA stimulus pattern. However, a paired t test indicated that there was a significant ($p < .02$) difference between RT(BBA) and RT(ABA). Like RT(BAA), RT(ABA) was practically equal to the overall mean RT, making no contribution to the repetition effect. The upper branch of the tree structure in Fig. 2 also has definite theoretical implications. On the basis of the trend displayed in the upper branch, it appears that the time required to process a given Signal A is directly dependent on the number of successive presentations of Signal B that have immediately preceded the presentation of Signal A. It would be interesting to compare the slope of this equation with those obtained in other experimental situations that employed two

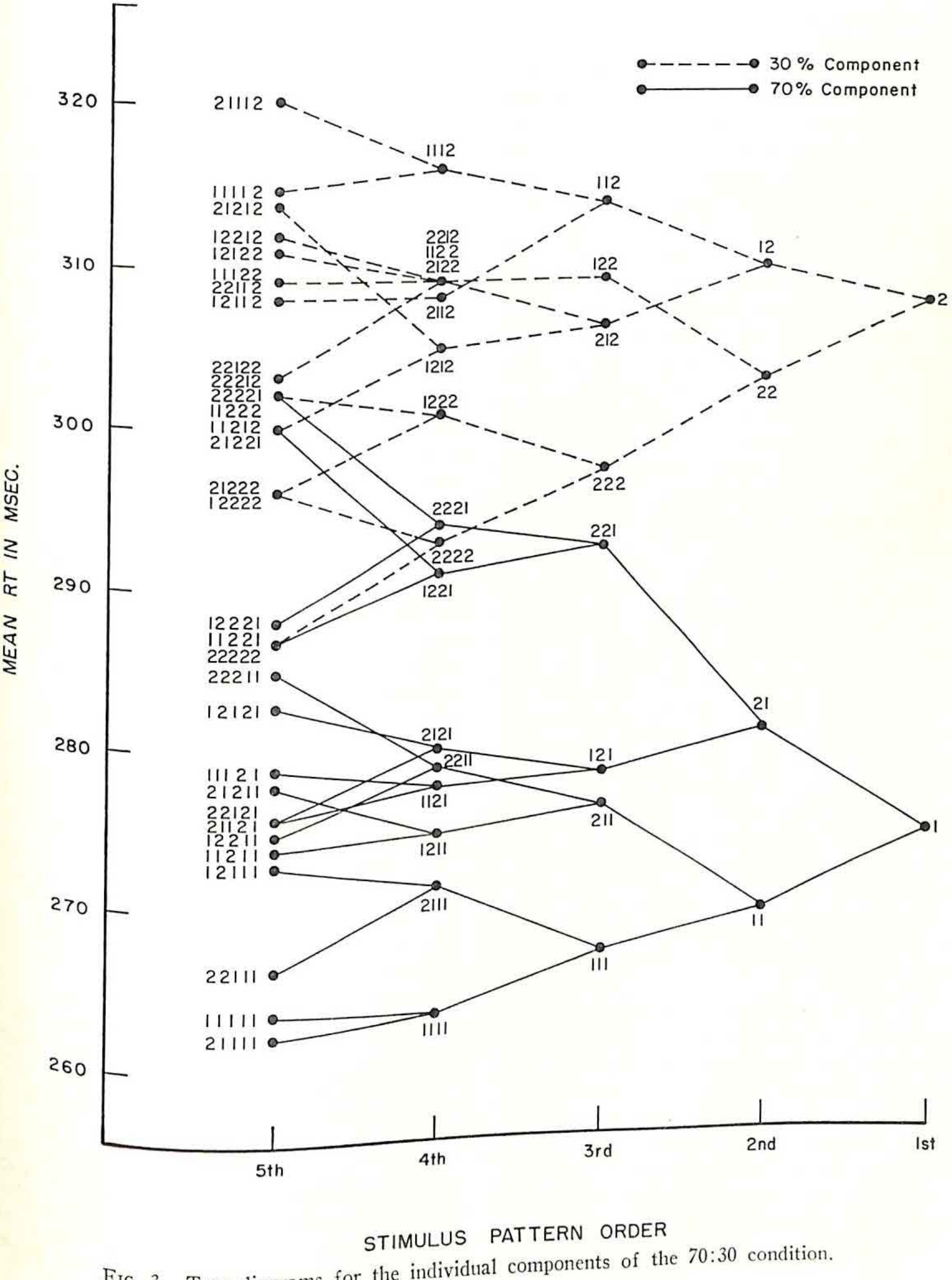


FIG. 3. Tree diagrams for the individual components of the 70:30 condition.

equiprobable signals. However, as far as the present author has been able to determine, previous researchers have not made this type of analysis. In fact, it appears that previous researchers have not examined and reported RT(BBA), RT(BBBA), or RT(BBBBA).

The results of the tree diagram analysis of the data from the 70:30 condition are presented in Fig. 3. In general, the results of these analyses support the major findings from the 50:50 condition regarding the importance of higher order stimulus patterns in determining RT and the nature of the repetition effect. For example, the results for both components provide added support for the contention that the repetition effect cannot be attributed to a single repetition of a stimulus event. Once again the observed repetition effect was mainly a result of the faster RTs associated with at least two repetitions (111 or 222, but not 211 or 122) and the slower RTs associated with stimulus patterns in which a stimulus was immediately preceded by at least two presentations of the alternate stimulus event (e.g., 112 or 221, but not 212 or 121).

The linear equations that best describe the trends for the *lower* branch of the tree diagrams (the repetition function) for each component presented in Fig. 3 appear to have different slopes. That is, the slope for the 30% component appears to be greater than the slope for the 70% component. These slopes were found for each *S*, and the differences were used in a paired *t* test. The results of this test showed that the slopes were significantly different ($p < .05$). This finding indicates that signal probability influences the amount of facilitation brought about by successive repetitions of the signal.

On the other hand, a similar comparison (paired *t* test) of the slopes for the linear equations for the *upper* branch of the tree diagrams showed that the slope for the 70% component was significantly ($p < .05$) larger than the slope for the 30% component. This finding indicates that successive presentations of the 30% stimulus immediately preceding the presentation of a 70% stimulus had a more detrimental effect on processing the 70% stimulus than vice versa. These findings illustrate the strength of the sequential effects and the type of phenomena that an adequate model of CRT should be able to predict.

The results of the present study, in addition to demonstrating the feasibility of a more comprehensive approach to the study of sequential effects, have clearly demonstrated that an adequate model of CRT must account for third, fourth, fifth, and possibly higher order sequential effects. Most of the current models of sequential effects in CRT are based on the theoretical notion that the time required to process a given signal is relatively independent of the signals presented more than one trial back (e.g., the model presented by Bertelson, 1963). The general structure of the tree diagrams reported in the present paper suggests that this theoretical notion is unrealistic. If stimulus patterns beyond the popular second order had no significant effect on RT, the tree structure resulting from a tree diagram analysis would not be expected to branch in any systematic fashion beyond the point representing the nodes for the second-order stimulus patterns. However, the resulting tree structures displayed systematic patterns of branching and nonnegligible sequential effects at every level examined. Furthermore, the resulting tree structures suggest that certain higher order sequential effects are as important as the more popular repetition effect (e.g., third-order sequential effects similar to the repetition effect in magnitude were found under both experimental conditions). Higher order sequential effects therefore deserve similar study and treatment in any theoretical formulation attempting to account for sequential effects in CRT.

It appears that premature theorizing based on the repetition effect has been relatively ineffective in producing models with adequate explanatory and predictive power. Further detailed data analysis of the type reported in the present paper should prove useful in sharpening present theoretical formulations. Similar data from a number of experimental situations would allow more precisely the specification of the requirements that an adequate theoretical model must satisfy. Fortunately, because of the theoretical work of Falmagne (1965), we already have a good example of some of the things that a model ought to predict. Falmagne's mathematical model is capable of making relatively good predictions concerning many interesting aspects of CRT data. However, there is convincing evidence that the processes underlying choice reaction behavior are too complex to be accurately represented by such a two-state model. The development of more sophisticated mathematical models would represent a challenging and perhaps rewarding approach. Finally, the type of detailed information that emerged from the tree diagram analyses introduced in the present study appears to be well suited to the computer simulation approach. Perhaps future work along these lines will yield a concise and comprehensive description of CRT behavior in the form of a computer program.

REFERENCES

Bertelson, P. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 1961, **13**, 90–102.

Bertelson, P. S-R relationships and reaction times to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, 1963, **65**, 478–484.

Bertelson, P. Serial choice reaction-time as a function of response versus signal-and-response repetition. *Nature*, 1965, **206**, 217–218.

Falmagne, J. C. Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology*, 1965, **12**, 77–124.

Hale, D. J. Sequential effects in a two-choice reaction task. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 133–141.

Hannes, M. The effect of stimulus repetitions and alternations on one-finger responding in two-choice reaction time. *Journal of Psychology*, 1968, **69**, 161–164.

Hyman, R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 1953, **45**, 188–196.

Kornblum, S. Choice reaction times for repetitions and non-repetitions: A re-examination of the information hypothesis. *Acta Psychologica*, 1967, **27**, 178–187.

Leonard, J. A., Newman, R. C., & Carpenter, A. On the handling of heavy bias in a self-paced task. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 130–141.

Smith, M. C. Repetition effect and short-term memory. *Journal of Experimental Psychology*, 1968, **77**, 435–439.

(Received February 20, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 258–262

IMPROVED RECALL FOR DIGITS WITH DELAYED RECALL CUES[1]

ROBERT G. CROWDER[2]

Yale University

Ten highly practiced *S*s recalled nine-digit series under five conditions of presentation. The control condition was compared with four conditions where the stimulus as presented included a redundant element at the end (i.e., a suffix). The delay between the last to-be-remembered element and the redundant suffix was .5, 2, 5, and 10 sec., respectively, in the four experimental conditions. The results showed that all conditions with the redundant suffix led to more errors than the control condition. However, only when the shortest suffix delay was used was there a selective impairment of performance in the last few serial positions. These findings were consistent with a set of hypotheses about storage in the auditory system and permitted estimation of the decay constant for such auditory storage to between 1.5 and 2 sec.

Crowder and Morton (1969) have recently presented evidence for a short-term sensory storage system specific to auditory information. Precategorical acoustic storage (PAS) was proposed by them to be a relatively peripheral and relatively transient buffer store in which only auditory information (i.e., only stimulus events which enter through *S*'s ears) is held prior to the process of recognition. It was maintained that the properties of PAS are qualitatively similar to those of the visual short-term sensory store (Averbach & Coriell, 1961; Sperling, 1963) but that PAS has a considerably longer storage duration. Information is lost from PAS for either or both of two reasons, according to the Crowder-Morton position: (*a*) a process of displacement or overwriting by subsequent auditory events, and (*b*) decay with time. One purpose of the present study was to estimate the time course of this decay.

The behavioral consequences of PAS are most readily observed in the stimulus suffix effect (Crowder, 1967; Crowder & Morton, 1969; Dallet, 1965; cf. Morton, 1968). The "suffix" is an extra digit, usually zero, read by *E* following presentation of the immediate memory series, just prior to *S*'s recall attempt. In studies of the stimulus suffix effect, *S* is informed in advance that over a block of trials each series will end with the same redundant digit and that he will not be responsible for recalling it. Since the redundant zero was never allowed to occur in the to-be-remembered series in these studies, it had the *logical* status of a recall cue, just as would the ringing of a bell following presentation of the memory series. The effect of a stimulus suffix (as compared with control series) is a sharp increase in errors localized mainly in the last few positions of the serial position function. The fact that the stimulus suffix effect is selective upon the terminal list positions led Crowder and Morton to attribute it to displacement of auditory information from PAS. (It is important to note that all of the studies cited so far using a stimulus suffix have employed *auditory* presentation of both the memory series and of the redundant suffix.) If the zero displaces information from PAS, it would not be expected to change error rates on early serial positions, because (*a*) information about the early positions would be too old to have survived in PAS until the suffix occurred, and (*b*) auditory presentation of the final list positions would have displaced PAS traces of the earlier elements. Thus, during presentation at conventional rates, each element has an overwriting effect on those immediately preceding it. The stimulus suffix effect shows that such displacement is inevitable, whether or not the displacing element carries information necessary for the task. In this regard, it is striking that a stimulus suffix reduces recall to the same degree as does an additional to-be-remembered element (Dallett, 1965). Dallett showed that recall for a seven-digit series which had been followed by a suffix was indistinguishable from recall for the first seven elements of an eight-digit series with no suffix.

---

[1] This research was supported by National Science Foundation Grant GB 4066. The careful assistance of Kathleen A. Catanese in data collection and tabulation is acknowledged.

[2] Requests for reprints should be sent to Robert G. Crowder, Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510.

In support of their interpretation of the stimulus suffix effect, Crowder and Morton (1969) showed that following *visual* presentation of the memory series, there was no selective effect of an *auditory* suffix at the end of the list. Thus, the effect was shown to be a strictly auditory phenomenon. The present experiment was designed to test other hypotheses about PAS through use of the suffix condition. If the effect of an extra digit is to displace information from PAS, and if information decays from PAS with the passage of (unfilled) time, then a sufficiently delayed suffix should not affect performance as does an immediate suffix (where "immediate" means a suffix presented in time with the rate of presentation of the memory series). Control (no suffix) conditions were therefore compared with conditions where the redundant suffix followed presentation of the memory series by varying amounts of time. The predictions were that performance would improve with longer series–suffix intervals, and that such improvement would be confined to the last few serial positions.

An experiment of Conrad (1960) provides an interesting comparison with the present study, for he too examined serial recall for digit lists as a function of the delay between stimulus presentation and the occurrence of a redundant digit (nought). In Conrad's study, however, the extra digit was supplied by *S* rather than by *E*. He had earlier shown (Conrad, 1958) that such a *response prefix* seriously impairs performance. The investigation in question (Conrad, 1960) revealed that delaying prefix emission by 5 or 10 sec. following presentation had no effect on performance. Nor did delaying the recall of unprefixed control series by the same amounts affect performance. In terms of their serial position effects, the response prefix and stimulus suffix effects have been shown elsewhere to be quite different (Crowder, 1967, Exp. III). Conrad's evidence holds that delaying the prefix has no effect on its impairment of recall. Thus, one additional distinction between the suffix and prefix effects (beyond the serial position difference mentioned previously) would be evidence from the present investigation that delaying the stimulus suffix *does* affect the magnitude of its effect on recall.

Another feature of the present design was the use of relatively few *S*s and several days of practice for each. If the PAS mechanism is indeed precategorical and unaffected by behavioral strategies, as Crowder and Morton (1969) maintained, then the stimulus suffix effect should persist even after large amounts of practice with suffix conditions.

METHOD

Design and materials.—There were five conditions, in all of which the to-be-recalled stimuli were nine-digit series taken from a table of permutations of the nine digits. In the control condition there was no redundant element; however, in each of the other conditions the digit "zero" was presented after the memory series. The delay between the last nonredundant digit and the zero was .5, 2, 5, or 10 sec., respectively, in the four experimental (suffix) conditions. Since the rate of presentation of the nonredundant elements was 2/sec, the .5-sec. delay was designated "immediate." All stimuli were recorded on magnetic tape, with 30 sec. allowed for recall (i.e., 30 sec. between presentation of the suffix digit and the beginning of the next series). The *S*s were 10 Yale undergraduates who volunteered to serve for pay. Each of the 10 *S*s served for 5 daily sessions with 15 trials apiece of the five conditions given each day. Motivation was supported by announcing cash awards (beyond the $1.50 rate per session) to the 3 *S*s who made fewest errors. A set of Latin squares was used to insure (*a*) independence of individual 9-digit stimuli from experimental conditions, (*b*) within-session balancing of practice effects across conditions, and (*c*) balancing of individual stimuli across *S*s and stages of practice.

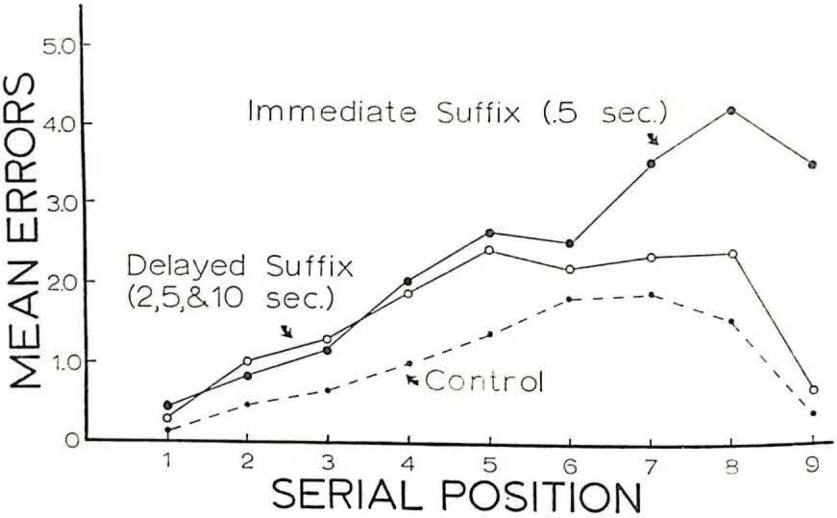


FIG. 1. The relation between mean error rate (maximum of 15) and serial position as a function of conditions. (The three delayed suffix conditions—2, 5, and 10 sec.—were combined.)

Procedure.—Each daily session contained 75 trials, divided into five blocks of 15 trials per condition. Following general instructions, there were 3-min. periods of rest before each block of trials and reminder instructions dealing with the procedure to be followed on the next block of trials. In the control condition, *S*s were told to recall the digit series out loud as soon as presentation terminated. Their recall was aided in all conditions by a position grid (Crowder, 1967) to insure useful serial position data. In the four suffix conditions, they were told to consider the digit zero as a cue to begin recall and, of course, not to emit the zero in recall. The period intervening between the last nonredundant element and the suffix was unfilled, and *S*s were given no rules as to what they could or could not do during this period.

RESULTS

The main results are given in Fig. 1, which shows serial position functions for the control, immediate (.5) suffix, and delayed (2, 5, and 10 sec.) conditions separately. The data of Fig. 1 show that whereas there was some decrement at all serial positions in the suffix conditions, only with the "immediate" suffix was there a selective decrement associated with the last few serial positions. An overall analysis of variance (5 Days × 5 Conditions × 9 Serial Positions) showed significant main effects of conditions, F (4, 36) = 9.00, $p < .01$, and serial position, F (8, 72) = 7.51, $p < .01$; otherwise, only the Conditions × Serial Position interaction was statistically significant, F (32, 288) = 7.03, $p < .01$. Inspection of the five conditions showed, however, that the three delayed-suffix conditions (2, 5, and 10 sec. between last digit and suffix) were virtually indistinguishable. A separate analysis of variance (5 Days × 3 Conditions × 9 Serial Positions) involving these delayed-suffix conditions only indicated that they were statistically as well as visually indistinguishable; the only statistically significant F was that for serial position, F (8, 72) = 8.44, $p < .01$. Therefore, the three delayed-suffix conditions were combined as they appear in Fig. 1 and submitted to an analysis of variance (5 Days × 3 Conditions × 9 Serial Positions). As before, conditions and serial position were statistically significant main effects, F (2, 18) = 16.49, and F (8, 72) = 7.05, respectively, p's both < .01, as was the interaction of these two factors, F (16, 144) = 10.88, $p < .01$. Practice (days) and its interactions with the other factors were nonsignificant.

Inspection of performance on the fifth day only indicated that the pattern of results in Fig. 1 persisted throughout the duration of the experiment. Mean error rates on the fifth day (maximum of 15) for Serial Positions 1–9 were: for the control conditions, .2, .7, .6, 1.3, 1.9, 2.0, 1.6, 1.5, and .6; for the immediate-suffix condition, .7, .8, 1.2, 1.8, 2.8, 3.1, 3.6, 3.9, and 4.3; and for the delayed-suffix conditions, .4, 1.2, 1.5, 2.4, 3.0, 2.5, 2.5, 2.5, and .8.

DISCUSSION

For the present purposes, the stimulus suffix effect may be defined as a selective performance decrement in the last few serial positions. Thus the effect was replicated only in the condition where .5 sec. intervened between the memory series and the redundant zero; at longer delays there was no *selective* error increase. It may be observed in Fig. 1 that there were *nonselective* increases in errors at all serial positions, including the early ones, for each of the five suffix conditions. Crowder and Morton (1969) have suggested that the nonselective effect of a suffix is due to articulatory interference, similar to that produced by a redundant response prefix. Strictly speaking, the present design does not permit a choice between such interference and a simple delay effect in accounting for the overall poorer performance of the conditions with 2-, 5-, and 10-sec. delayed suffixes as compared with controls. However, Conrad's (1960) data showed that such a delay, without redundant interpolated material, had no significant effect on performance. Conrad's data also indicate that the difference between immediate- and delayed-suffix conditions in the present study was not due to the passage of time alone, although time was obviously confounded with suffix delay in the experimental design. (Even if Conrad's results were not available, the retention function for nonsuffix control series would have to be most eccentric to account for the present results.) The important finding of the present study was that the characteristic skewing effect of a stimulus suffix on the shape of the serial position curve was obtained only when the suffix followed the last memory element by .5 sec. Moreover, the present data indicate that the effects of such an immediate suffix remain at least as dramatic after 60 trials of that condition (and 180 delayed-suffix trials) as they were initially. Apparently, even though the logical requirements of the immediate-suffix task permit *S* to completely ignore the suffix digit, he cannot do so.

Technically, the improved recall obtained when the suffix was delayed as opposed to immediate qualifies as "reminiscence," although such designation does not in the slightest illuminate the underlying mechanism. It is the selectivity of this "reminiscence" in the last few serial positions that has theoretical importance. Every one of the 10 *S*s showed more errors in both Positions 8 and 9 with the immediate than the delayed suffix; in Position 7, 9 out of the 10 *S*s showed the same difference. These data are consistent with predictions from the PAS hypothesis. Moreover, the fact that performance was independent of suffix delay after 2 sec. indicates that the duration of information in PAS is necessarily shorter than 2 sec.; otherwise, there would have been some effect of the 2-sec. suffix on recall for at least the ninth serial position. Thus, an upper bound on the limit of PAS storage is 2 sec.

Looking at the data of Fig. 1 another way allows a second estimate of PAS duration. The "immediate" suffix occurred .5 sec. after the ninth digit. A lower bound estimate of PAS duration would be indicated by the "distance back" into the series over which the .5-sec. suffix was effective. The proper base line for making this comparison is the combined delayed-suffix conditions since they share with the immediate conditions whatever nonselective effects result from suffix presentation. The curves of Fig. 1 show that up through the sixth digit, it made no difference whether the zero was presented immediately or delayed; however, for the last three positions there was an increasing degree of impairment by the immediate suffix. Therefore, a tentative lower bound estimate of PAS duration is 1.5 sec., since this was the period of time elapsing between the seventh element and the suffix.

Although these two estimates of PAS decay characteristics are in remarkable agreement, no particular generality is claimed for them beyond the exact procedures of the present study. Differences in loudness, intelligibility, and other factors may prove in future work to be related to the duration of storage in the auditory system.

The dependence of the stimulus suffix effect on the delay between stimulus presentation and occurrence of the redundant element demonstrated in the present study, is in marked contrast to the independence of the response prefix effect on such delays (Conrad, 1960). Further evidence is therefore adduced that whether *E* or *S* emits a redundant element between presentation and recall makes a substantial difference.

REFERENCES

AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell System Technical Journal*, 1961, 40, 309–328.

CONRAD, R. Accuracy of recall using keyset and telephone dial, and the effects of a prefix digit. *Journal of Applied Psychology,* 1958, 42, 285-288.

CONRAD, R. Very brief delay of immediate memory. *Quarterly Journal of Experimental Psychology,* 1960, 12, 45–47.

CROWDER, R. G. Prefix effects in immediate memory. *Canadian Journal of Psychology,* 1967, 21, 450–461.

CROWDER, R. G., & MORTON, J. Precategorical acoustic storage (PAS). *Perception and Psychophysics,* 1969, **5**, 365–373.

DALLETT, K. "Primary memory": The effects of redundancy upon digit repetition. *Psychonomic Science,* 1965, 3, 237–238.

MORTON, J. Selective interference in immediate recall. *Psychonomic Science,* 1968, **12**, 75–76.

SPERLING, G. A model for visual memory tasks. *Human Factors,* 1963, **5**, 19–31.

(Received February 20, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 263–266

RETRIEVAL OF COLOR INFORMATION FROM PREPERCEPTUAL MEMORY[1]

SANDRA E. CLARK[2]

Ohio State University

Two experiments were conducted to investigate preperceptual memory after presentation of an array of colored stimuli. The *S*s in each experiment viewed an array of red, green, and yellow circles. At predetermined intervals after termination of the display (0–1,250 msec.), a tone retrieval cue indicated the type of color information to be reported, either a partial report (one of the three colors in Exp. I, colors in one of the three rows in Exp. II) or a complete report. In Exp. I, color information reported by category did not decrease in availability as delay of retrieval increased; however, Exp. II indicated that availability of color location information did decrease. These results imply a differential processing strategy depending on the type of information to be retrieved.

Information within an organism can be described according to the flow model of Broadbent (1963), which assumes a variety of information states between the input of the proximal stimulus and the motor or verbal output. Accordingly, visual information may be held in the form of a raw neural message before categorization or may be held in a state resulting from categorization. Melton (1963) has described preperceptual information as a visual stimulus not encoded with central or peripheral response components, while postperceptual information is a memory trace resulting from central encoding or analysis.

A model for the two storage systems has been proposed by Broadbent (1958), who advocated a storage system for sensory data (the S system) and a separate system (the P system) for analyzed information. According to Broadbent, sensory messages are dealt with in two stages: (*a*) The general physical features of the stimulus are identified, in which several messages may be handled at once; (*b*) the verbal content is analyzed, in which only one message can be handled. Several studies (Treisman, 1964; Turvey, 1966, 1967) have shown evidence for these two operationally different stages.

In vision, the main physical dimensions of stimuli are location, color, size, brightness, and shape. Sperling (1960) and Averbach and Coriell (1961) have shown that the availability of briefly presented visual information is a decreasing function of the delay between presentation and retrieval. They also showed that accuracy of a partial report is greater than that of a complete report. These results indicate that location of letters is identified at the first stage of information processing. However, Sperling showed that no more information was given in a partial report of letters from a letter–number array than in a complete report of the entire display, indicating that categorization of items takes place in the second stage.

Treisman (1966) stated the need to show that *S*s can pick out visual stimuli based on other physical dimensions other than location. While working with a two-color, letter–number array, Dick (1967) found a significant effect of delay in a partial report of location and of color, but no effect for identity.

The present experiments investigated further the partial report of color and location from an array of colors that has been briefly displayed.

---

[1] This study is based on a dissertation submitted to Ohio State University in partial fulfillment of the requirements for the PhD degree. The research was supported in part by National Institute of Mental Health Grant MH-8526 to Delos D. Wickens, who supervised the thesis.

[2] Requests for reprints should be sent to Sandra E. Clark, who is now at the University of Wisconsin–Milwaukee, Department of Educational Psychology, Milwaukee, Wisconsin 53201.

METHOD

Experiment I

Subjects.—The five *S*s were volunteers from introductory psychology classes who received extra-course credit.

Materials.—The *S*s were presented a slide containing 15 colored circles (equal numbers of red, R; green, G; and yellow, Y) randomly arranged in three rows of 5 circles. The information was displayed onto a dark preexposure field containing a fixation point maintained by a Revere 888 slide projector. A Carousel 800 slide projector with an Alphax-Wollensak tachistoscope shutter displayed the information for 150-msec. duration onto a screen. The slide subtended approximately 5° of visual angle to *S*. The postdisplay cue followed at delays of 0, 100, 250, 500, 750, 1,000, and 1,250 msec. as timed by Tektronix pulse generators. The postdisplay cue was one of four tones (sine wave form) generated by an Eico audio-generator and presented via earmuff headphones to *S*.

Procedure.—On Day 1, *S* was told that the study was testing his ability to retain a visual image and was then shown a sample slide. The *S* was oriented to the procedure and to the four different auditory poststimulus cues, which signaled the color to be reported or a complete report. The tones were adjusted individually for each *S* to obtain maximum ease of differentiation.

The procedure for stimulus display involved *S* pressing a key to view a slide after *E*'s statement of "Ready." The *S* was told to look at the fixation point until it was clearly in focus and then to press the key, which triggered the shutter on the Carousel projector. This allowed *S* to be prepared for the visual presentation. The first three poststimulus tones required that *S* report the locations of one of the three colors by marking the letter of that color (R, G, or Y) in the appropriate spaces of an answer grid of three rows with five squares in each row, a square corresponding to each circle on the stimulus slide. The fourth poststimulus tone required *S* to report all colors by R, G, and Y in all squares of the answer grid. The *S* was encouraged to always fill in as many squares as possible, but not to "guess wildly." As soon as *S* completed his answer, the grid was passed to *E* so that it was no longer available to *S*. The interstimulus interval was never less than 10 sec. or longer than 15 sec. The *S* received 25 practice slides.

On Days 2 and 3, 80 slides were presented (40 for complete report and 40 for partial report). The partial-report slides had all colors reported, and both types of report occurred at all retention intervals. The type of report, color reported, and retention intervals were randomized. The order of presentation of the slides was counterbalanced across *S*s.

The purpose of the preexperimental procedure was to insure that (*a*) *S* had associated the poststimulus cues with the respective tasks with at least 95% accuracy, (*b*) *S* was demonstrating equal attention (all rows and all colors were equally available), and (*c*) practice effects would be minimal on the experiment days (Days 4 and 5).

On Days 4 and 5, 72 slides were presented (36 partial and 36 complete report). Each type of report and each color in the partial report was used equally often at each retention interval. Retention intervals were randomized within blocks of 36 trials. A set of slides different from those of the preexperiment days was used. The order of presentation of slides was counterbalanced across *S*s.

If *S* misinterpreted a tone and wrote either the wrong color or made an incorrect type of report, the slide was shown again after the 72 trials for the experimental days had been completed. That procedure was possible since Turvey (1967) has shown no effect of repetition in terms of percentage of recall with an unanalyzed stimulus.

Treatment of the data.—A measure was necessary to enable comparison between the partial and the complete report. According to Sperling (1960), the partial report is considered as a random sample of the information that *S* has available for report. For example, if *S* is correct 80% of the time when he is reporting 5 of the 15 color circles, then he is said to have 80% of the 15 colored circles (12 colored circles) available for partial report at the time the retrieval tone is presented.

In order to calculate the number of available colored circles, the number of colored circles correct in the partial report is multiplied by the number of equiprobable partial reports. With three colors, multiplication is by 3.0.

Sperling (1960) has shown that the amount of information available to *S*, calculated from the partial report, is greater than that given in the complete report, which is invariant with delay of poststimulus cueing. Thus, subtracting the number of correct responses in the complete report from the number of responses shown available by partial report would indicate the additional number of responses available to *S*, excluding the immediate memory span. For example, if nine responses were calculated to be available and the number of correct responses in the complete report was five, *S* would have four additional responses available.

Experiment II

Subjects.—The five *S*s were volunteers from introductory psychology classes who received extra-course credit.

Materials and procedure.—The materials, procedure, and data treatment were similar to that of Exp. I except that the first three retrieval cue tones required that *S* report the colors located in one of the three rows of the stimulus. The fourth retrieval cue tone required *S* to report all

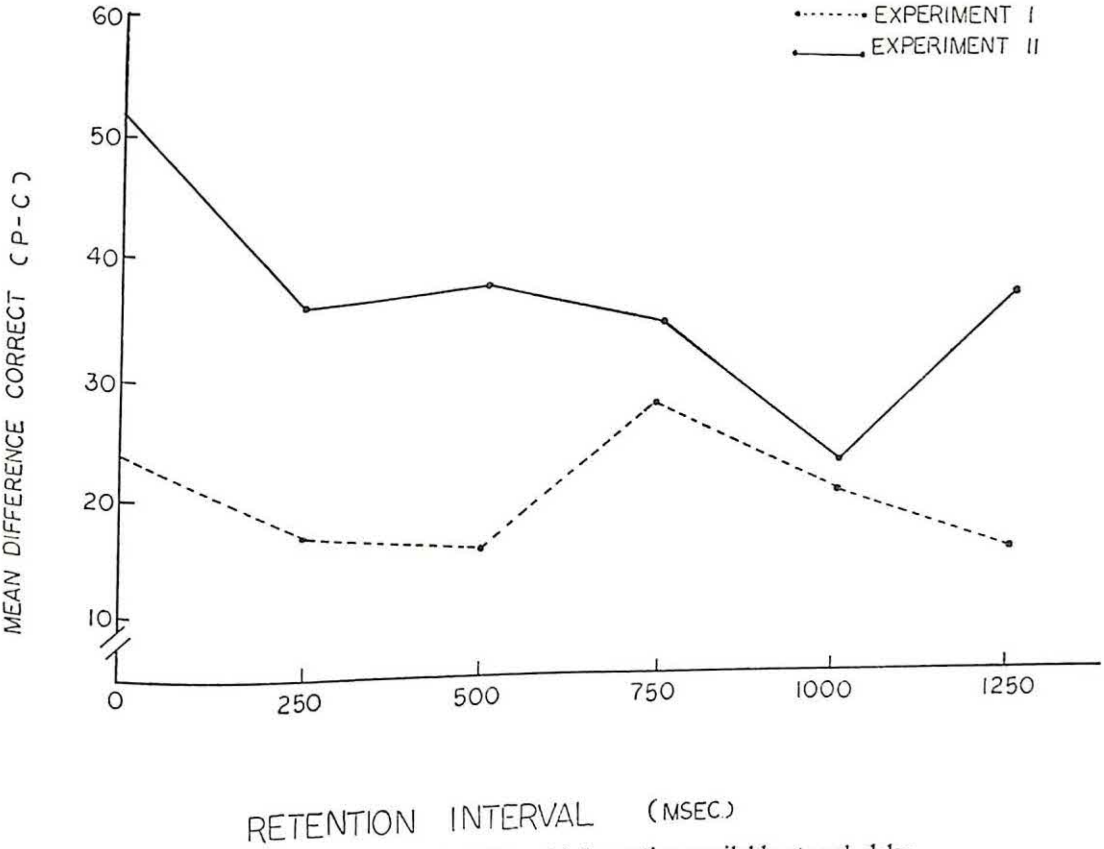


FIG. 1. Mean amount of additional information available at each delay interval for each experiment.

rows in the stimulus. Time intervals, type of report, and row reported were randomized within blocks of 36 trials.

RESULTS

The mean additional information available (mean calculated partial-report score, P, minus the mean number of complete-report responses, C) at each delay interval for Exp. I and II is shown in Fig. 1. A repeated-measures analysis of variance on the partial–complete scores indicated a significant effect of delay, $F\ (5, 40) = 3.43$, $p < .01$, and a significant Experiment × Delay interaction, $F\ (5, 40) = 3.16$, $p < .05$. The effect of experiment was not significant, $F\ (1, 8) = 4.93$.

Since the Cue Delay × Experiment interaction was significant, simple F tests were performed on the main effect of experiment. The F test for Exp. I indicated no difference in cue delay, $F\ (5, 40) = 1.33$. The F test for Exp. II indicated a significant difference in cue delay, $F\ (5, 40) = 5.26$, $p < .01$. A Newman-Keuls test was performed on Exp. II and indicated significant differences in retention between 0 delay and each of the other delay intervals ($p < .05$ for all delays except 1,000 msec., $p < .01$). No other delay intervals were significantly different.

DISCUSSION

Experiment I indicated that *S*s were able to report by category of information (a particular color). These results differ from those of Sperling (1960), who found with a mixed array of letters and numbers that there was no more information available for either letters or numbers than that contained in the immediate memory report. Thus, when information can be immediately encoded neurally on stimulus presentation (as with color), *S*s can report by category; however, when information cannot be immediately encoded (as with letters or numbers), *S*s cannot report by category of information. The results of Dick (1967) with color information also support the present conclusions.

Further, in contrast to Sperling (1960), there was no decrease in additional color information available as the delay of the poststimulus cue was increased. The amount of additional color information available in Exp. I was invariant with time of delay.

However, in Exp. II there was a decrease in the amount of additional spatial location information as delay of retrieval increased. These results are similar to those of Sperling (1960), although he reported that after 1,000 msec., information available by partial report was no different from that of the immediate memory span. In Exp. II, after the same time interval, considerable information was still available. Thus, when color information is reported by spatial location rather than by category, there is a decrease in availability of that information as delay of retrieval is increased.

The difference between the results of Sperling (1960) (no more information available by partial report when using categories) and those of the present study are consistent with the model of information processing as proposed by Broadbent (1958). The first stage of processing deals only with the general physical features of the stimulus. When color categories are employed, analysis is possible during Stage 1, and thus more information is available using the partial report. However, when verbal content must be analyzed, as with a mixed letter–number array, analysis does not take place during the preperceptual Stage 1. Hence, no more information is available with a partial report than with the immediate memory span.

When color information is presented, availability of category information does not decrease during retrieval delay, but the availability of spatial location information does decrease. Dick (1967) has suggested that a decrease in accuracy may be the result of sequential processing of information, while no decrease may be due to a parallel processing mechanism. It is possible that different encoding strategies are employed due to the type of information retrieval that will be required. When color category information must be retrieved, color is quickly encoded with location in a parallel fashion and thus protected from decay. When location information must be retrieved, location and color are sequentially encoded. Due to the longer time needed for encoding, the information is more subject to decay. Dick (1967) has found similar results with location and identity.

Thus, any preperceptual memory model must systematically describe the length of time for neural encoding to occur. The stage model of information processing suggests that the results of the present experiments could be applicable to the other information of the physical attributes of a stimulus, i.e., shape, size, and brightness.

REFERENCES

Averbach, E., & Coriell, A. S. Short-term memory in vision. *Bell Systems Technical Journal,* 1961, **40**, 309–328.

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Broadbent, D. E. Flow of information within the organism. *Journal of Verbal Learning and Verbal Behavior,* 1963, 2, 34–39.

Dick, A. O. Short-term memory for several stimulus attributes. Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1967.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior,* 1963, 2, 1–21.

Sperling, G. The information available in brief visual presentations. *Psychological Monographs,* 1960, **74** (11, Whole No. 498).

Treisman, A. Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3**, 449–459.

Treisman, A. Human attention. In B. M. Foss (Ed.), *New horizons in psychology.* Baltimore: Penguin, 1966.

Turvey, M. T. The effects of rehearsing analyzed information upon the retrieval of unanalyzed information. *Psychonomic Science,* 1966, **6**, 365–366.

Turvey, M. T. Repetition and the preperceptual information store. *Journal of Experimental Psychology,* 1967, **74**, 289–293.

(Received February 24, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 267–271

SOME EFFECTS OF SIMULTANEOUS FORCE-PROPORTIONAL POSITIVE AND NEGATIVE REINFORCEMENT[1]

ROSS D. L. FILION,[2] STEPHEN C. FOWLER, AND J. M. NOTTERMAN[3]

Princeton University

Two groups of rats were exposed to the same force-proportional positive reinforcement schedule, but under different schedules of simultaneous force-proportional negative reinforcement. The data indicate that force of response and rate of responding are differentially sensitive to punishment and that the nature of this effect is such as to suggest the existence of a tendency for compensatory interaction between the two dependent variables.

The experiment here reported serves to augment an earlier account of force-proportional reinforcement, one which was concerned solely with an examination of positive reinforcement contingencies (Notterman & Mintz, 1965). It does so by making not only amount of food, but probability of receiving shock as well, each stimultaneously proportional to the peak force of individual bar-pressing responses.

The principal finding of the previous research was that *S*s (laboratory rats) did not maximize the opportunity to press "harder for more." It was concluded that the organism reached an upper limit of optional exertion which probably depended on the interaction of several variables: (*a*) the absolute amount of reinforcement obtained per response, as it influenced drive state; (*b*) the exertion required to obtain this amount, the value of "gain" being set by *E*; (*c*) the choice available of pressing again, rather than harder; and (*d*) the capacity of *S* to discriminate changes in drive cues, amount-of-food cues, and in motor feedback cues.

Other things being equal, and depending on the particular intensity of shock used, it might be expected that the inclusion of punishment as an additional variable should serve to decrease overall rate of responding (Church, 1963; Estes & Skinner, 1941; Solomon, 1966). However, no clear evidence is available to permit prediction of the effect of shock on level of response force, especially in a force-proportional situation. For example, if it is assumed that the rate-depressing effect of shock extends to force of response, then the consequence of including probability of receiving shock in a force-proportional reinforcement contingency should be such as to induce not only lower rate, but also lower peak force. On the other hand, it might be hypothesized that the greater the penalty (or probability of receiving shock) for pressing harder during an individual response, the greater the tendency for *S* to choose to press less vigorously, but more often. This tendency, should it indeed exist, might be sufficient to counter the previously reported decreases in rate. Moreover, it might lend support to arguments for the descriptive utility of concepts such as "reinforcement earned per minute," since the options available to *S* in this experiment permit just such a unifying principle to become operative (Herrnstein, 1961; Neuringer, 1967; Skinner, 1966). This interpretation would require the further assumption that the values of shock intensity, shock probability, and positive reinforcement gain have been appropriately chosen.

In sum, the experiment was designed to provide information related to the following issues: (*a*) Given the same force-proportional, positive reinforcement gain, but under two values (low or high) of simultaneous, negative reinforcement gain, how are

---

[1] This research was supported at Princeton University by the Office of Naval Research, Physiological Psychology Branch, under Contract NONR 1858(19).

[2] Now at the University of Pennsylvania Medical School, Department of Psychiatry.

[3] Requests for reprints should be sent to J. M. Notterman, Department of Psychology, Green Hall, Princeton University, Princeton, New Jersey 08540.

TABLE 1

FOOD AND SHOCK CONTINGENCIES AS A FUNCTION OF REQUIRED RESPONSE FORCE

Phase	Required in peak force (gm.)	Amount food obtainable/ response (No. pellets × 45 mg.)	Probability of shock	
			Group I	Group II
1 (Tray-approach)	None	45 mg.	0	0
2 (Shaping & CRF)	8–up	45 mg.	0	0
3 (Proportional reinforcement)	0.0–7.9	0 mg.	0	0
	8.0–15.9	45 mg.	.10	.20
	16.0–23.9	90 mg.	.20	.40
	24.0–31.9	135 mg.	.30	.60
	32.0–39.9	180 mg.	.40	.80
	40.0–up	225 mg.	.50	1.00
4 (Extinction)	None	0	0	0

peak force of response and rate of responding affected? Are these dependent variables equally sensitive to the risks incurred in receiving either low or high schedules of increasing shock probabilities? (*b*) Under the specific parametric conditions of the present experiment, are such changes in rate and in force as are observed so related as to provide support for "intake per minute" types of behavioral descriptions?

METHOD

Apparatus.—The apparatus was a modification of the analog computer system described by Notterman and Mintz (1965); instead of a strain gauge, a Sanborn transducer (Model FTA 100-592) served as the manipulandum. The portion of the transducer available to *S* was a 13-mm.-diameter disc, the surface of which was 62 mm. above the grid floor of the Skinner box. The manipulandum was positioned outside the cage, but within reach of the animal's paw. Since excursion of the disc was less than .4 mm. for forces up to 200 gm., the manipulandum was essentially isometric. A Hewlett-Packard digital recorder (Model 561B) registered peak force (F_p) for all pressures exceeding 2.5 gm. (termed the "threshold"), which defined a response. Rate of bar-pressing and mean peak force of response were computed from these pressures.

Subjects and procedure.—The *S*s, 14 male Sprague-Dawley rats weighing 125–200 gm. upon receipt, were maintained on 22-hr. hunger drive throughout the experiment, being allowed free feeding of Purina Lab Chow mash for 1 hr., approximately 1 hr. after each session. At the outset of the experiment, they were randomly assigned to two groups of equal number and subsequently were exposed to three sessions of tray-approach training (Phase 1); four sessions of shaping based on continuous reinforcement (Phase 2); 20 sessions of proportional reinforcement (Phase 3); and three sessions of extinction (Phase 4). For pressures 8.0 gm. (termed the "criterion") and above, food and shock were dispensed in accordance with Table 1, but only after the response had terminated (i.e., emitted force fell below threshold). The sessions of the first three phases each were concluded on *S*'s obtaining approximately 40 45-mg. pellets; the sessions of Phase 4 were each 10 min. in duration, with *S* receiving neither food nor shock.

It is evident from Table 1 that during Phase 3, the specified gain (defined as the amount of reinforcement earned for a given range of emitted forces, divided by the lower limit of that force range) was 5.62 mg. of food/gm of response force for criterion forces and above. Shock level during this phase was adjusted for individual *S*s until the tenth session, by which time it was 80–85 v. ac for all *S*s. Scrambled shock was administered on a variable ratio basis through the grid floor, each force step being independently programmed by means of a Scientific Prototype Model 4020-J probability generator.

RESULTS AND DISCUSSION

Figure 1 shows the mean F_p for the separate groups over the course of proportional reinforcement. Considering this phase of the experiment in its entirety, Groups I and II are significantly different at beyond the .01 level, $F\ (1, 12) = 10.03$.[4] Moreover, *t* test indicates that the two groups differ on the very first session ($p < .01$) and that the effects of subsequent, successive sessions on both groups are negligible, $F\ (19, 228) = 1.60$, $p > .10$.[5]

The separation in force magnitude between Groups I and II cannot be attributed to differential responsiveness to the positive aspects of the force-proportional contingency for the following reasons. Apart from the fact that both groups were exposed to the same positive reinforcement regimen, it will be noted that the magnitude of mean F_p for both groups is well within or below the first force step (8.0–15.9) on each session—no upwards shaping of forces is detectable. Analysis indicates that the dif-

---

[4] Two factors design with repeated measures on the successive sessions or training factor (cf. Winer, 1962, p. 302.).

[5] The degrees of freedom are derived for the sessions factor by $(N-1)$, where $N = 20$, and for the successive mean F_p factor by $2(n-1)(p-1)$, where $n = 7$ and $p = 20$.

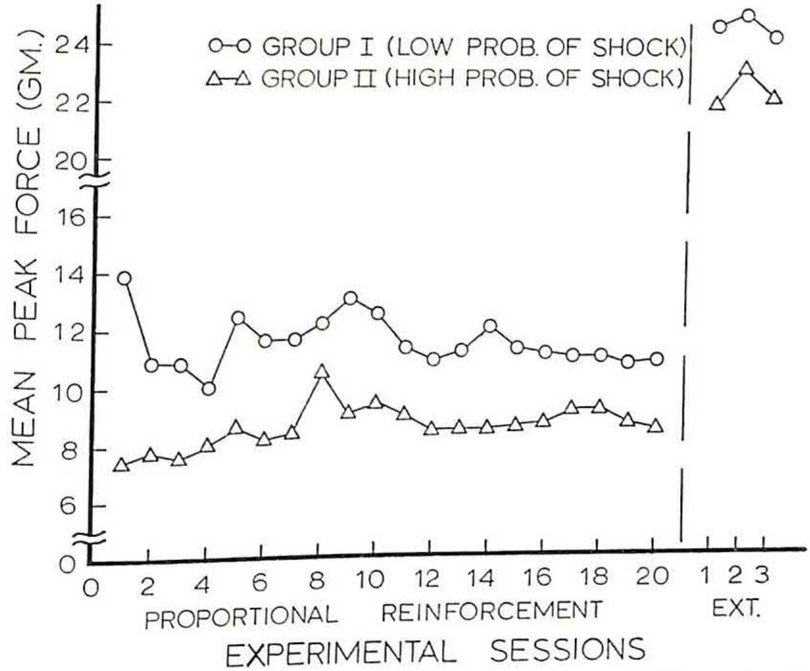


FIG. 1. Mean peak force of response as a function of acquisition and extinction. (Each point is a group mean based on seven *S*s.)

ference in mean F_p between Group I and II is probably accounted for by the greater frequency of subcriterion responses (i.e., pressures below 8.0 gm.) characteristic of the performance of Group II. The difference in relative frequency of subcriterion

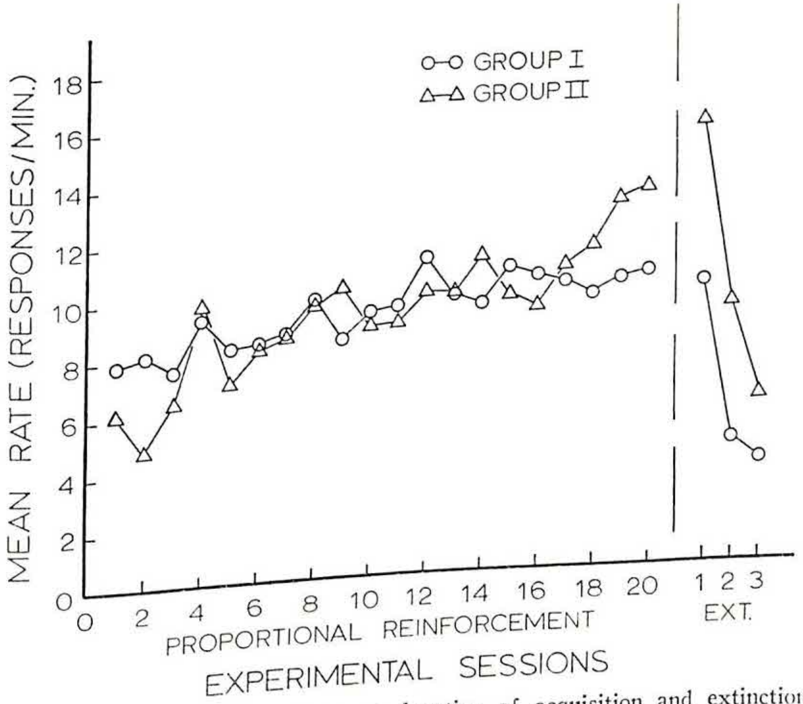


FIG. 2. Mean rate of response as a function of acquisition and extinction. (Each point is a group mean based on seven *S*s.)

responses, however, is not quite significant, $F\ (1,\ 12) = 3.86,\ .05 < p < .10$. Further support for the interpretation that the separation between the groups is representative of the greater suppressive effect of the higher shock probability to which Group II was exposed is afforded by comparing the mean F_p of Group I with that of another group which received no shock at all. A group of seven *S*s of similar strain and comparable weight upon receipt, but taken from another experiment, serves to provide a modicum of information on this point. Conditions for this third sample were similar to those of Groups I and II, except that there were 15 (as opposed to 20) sessions of proportional positive reinforcement during Phase 3. An analysis of variance comparing Group I and the no-shock group for the first 15 days of their respective acquisition phases reveals that there is no significant difference in mean F_p, $F\ (1,\ 12) = .07$, $p > .25$, nor is there a significant training effect, $F\ (14,\ 168) = 1.108$, $p > .25$. As would be expected, however, the difference between Group II and the no-shock sample over the same number of sessions is reliable, $F\ (1,\ 12) = 6.248$, $p < .05$.

As they enter into extinction, Groups I and II each increase in mean F_p. Comparing the last session of proportional reinforcement with the first session of extinction, the rise in force is significant by signed rank test for each group at the .02 level, every *S* increasing in force.[6] This finding is in accord with earlier reports (see Notterman & Mintz, 1965). However, the groups are not reliably different from each other, $F\ (1,\ 12) < 1$; nor is there a significant drop in mean F_p over the course of the three extinction sessions, $F\ (2,\ 24) < 1$.

The rate data depicted in Fig. 2 are noteworthy in that statistical tests show reliable effects which are opposite to those for the mean F_p measure. Unlike the case with mean F_p, Groups I and II do not differ in rate during proportional reinforcement, $F\ (1,\ 12) < 1$; on the other hand, the effect of training on both groups taken together is significant, $F\ (19,\ 228) = 4.43$, $p < .01$. Nor is the latter result compromised by the interaction term, $F\ (19,\ 228) = 1.29$, $p > .25$. In terms of rate, the no-shock sample is not significantly different from either Group I or II.

The fact that higher probability of shock had a significant effect on mean F_p but not on rate suggests that mean F_p may be more sensitive to differences between punishment procedures than is rate of responding. However, the disparity between the two measures during the course of Phase 3 may simply be a reflection of the particular shock intensity and probability used—higher intensities or probabilities might have suppressed both measures, whereas lower magnitudes might have suppressed neither.

Again in contrast to the mean F_p measure, the separation of the rate curves during extinction is reliable, $F\ (1,\ 12) = 11.10$, $p < .01$, and the decrease in rate with successive extinction periods is also significant, $F\ (2,\ 24) = 44.46$, $p < .001$. The former result might have been expected in view of the previously cited data concerning the relative frequency of subcriterion responses. If subcriterion responses are treated as the unreinforced responses of a partial reinforcement schedule, then it is possible to compute the value of the variable ratio (VR) schedule that each group imposed upon itself. For example, if one considers the last four sessions of Phase 3, then Group I had placed itself on a VR 1.3, but Group II had elected to respond on a VR 1.7. When examined in terms of intermittent schedules of reinforcement, the rate data for the extinction sessions are not so surprising: Group II was on a "leaner" schedule and was more resistant to extinction, a typical consequence of partial reinforcement.

In addition to providing evidence which suggests that punishment of a response can suppress the force of that response without abolishing responding, the data reveal a marked difference between F_p and rate measures. The early separation of the F_p curves and the lack of a detectable training effect imply that learning of the force requirement is very rapid, perhaps occurring within a few trials. Once a given level of performance in terms of mean F_p is established, it remains rather stable until some major adjustment of the reinforcement condition is made, e.g., extinction. On the other hand, rate measures seem to change much more gradually, and in the present study the positive slope of the rate curves in Phase 3 may be partially a result of adaptation to shock.

---

[6] A nonparametric test was used since the variance in extinction is quite higher than that of acquisition.

The existence of the slope in the rate curve can also be interpreted as being the consequence of a tendency for *S*s to obtain increasing amounts of reinforcement per minute as a function of training, independently of, or along with, adaptation to shock. This measure does, indeed, increase with sessions, going from an initial value of 392 to 414 mg/min for Group I, and from 252 to 345 mg/min for Group II. The upwards trend for both groups taken together is significant, F (19, 228) = 3.734, $p < .01$. The intergroup difference itself, however, is not quite significant, F (1, 12) = 3.86, $.05 < p < .10$; nor is there any proclivity shown for Group II to converge on Group I, the interaction term being unreliable, F (19, 228) < 1.

In answer to the questions posed earlier, it appears that force of response and rate of responding are differentially sensitive to the effects of increasing shock probabilities, at least under the conditions of the present experiment. There is also some supportative evidence for the utility of "reinforcement per minute" as a descriptive construct, but—in the absence of parametric variation—the evidence must be considered as only suggestive.

REFERENCES

Church, R. M. The varied effects of punishment of behavior. *Psychological Review,* 1963, **70,** 369–402.

Estes, W. K., & Skinner, B. F. Some quantitative properties of anxiety. *Journal of Experimental Psychology,* 1941, **29,** 390–400.

Herrnstein, R. J. Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior,* 1961, **4,** 267–272.

Neuringer, A. J. Choice and rate of responding in the pigeon. Unpublished doctoral dissertation, Harvard University, 1967.

Notterman, J. M., & Mintz, D. E. *Dynamics of response.* New York: Wiley, 1965.

Skinner, B. F. Operant behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application.* New York: Appleton-Century-Crofts, 1966.

Solomon, R. L. Punishment. In R. N. Haber (Ed.), *Current research in motivation.* New York: Holt, Rinehart & Winston, 1966.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received February 24, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 272–278

DIFFERENTIAL CLASSICAL EYELID CONDITIONING AS A FUNCTION OF CS INTENSITY, CS RISE TIME, AND INTERSTIMULUS INTERVAL[1]

SUSAN M. WILCOX AND LEONARD E. ROSS[2]

University of Wisconsin

Two experiments were conducted to investigate the effects of auditory CS properties on differential classical eyelid conditioning. The first experiment included two levels of background intensity (65 and 80 db.), two levels of CS intensity relative to background intensity, and interstimulus intervals (ISIs) of 500 and 800 msec. The ISI × CS Intensity, relative to background intensity, interaction was significant, but no effect was observed for either background intensity or absolute CS intensity. In the second experiment, CS rise time and ISI were varied, with the slower CS onset resulting in better differential performance than fast CS onset under shorter ISI conditions. Variations in CS rise time had no effect on differential performance with the longer ISI. Results were discussed in terms of a trade-off between ISI and the characteristics of the CS, with the longer ISI leading to better differential performance under CS conditions that were nonoptimal for the shorter ISI.

Although recent interest in the perceptual and/or inhibitory processes involved in differential eyelid conditioning has led to a number of studies investigating factors affecting differential conditioning, one variable, CS intensity, has received little attention. Evidence suggesting that within-*S* designs are necessary to obtain CS intensity differences in simple conditioning may be responsible for the neglect of this variable since investigation of within-*S* CS intensity effects in differential conditioning would require a relatively complicated design involving more than one cue dimension.

There are data, however, that indicate that CS intensity effects can be obtained in simple conditioning situations with between-*S* comparisons and therefore that between-S differential conditioning might also be related to this variable. Although early investigations of CS intensity in simple eyelid conditioning found no significant between-*S*s effect (Carter, 1941; Grant & Schneider, 1948), later studies have found evidence for higher levels of performance with a more intense CS (Mattson & Moore, 1964; Walker, 1960). A major difference between these two pairs of experiments was the interstimulus interval (ISI) used. In the Carter and Grant and Schneider studies, the ISIs were 760 and 750 msec., respectively, while the ISI employed in the two more recent studies was 500 msec. Since the optimal ISI for single-cue conditioning is generally believed to be approximately 500 msec., these results indicate that CS intensity effects are more likely to be obtained when the optimal ISI is used, and, in fact, a CS Intensity × ISI interaction has been obtained for simple acquisition in a within-*S*s design (Lipkin & Moore, 1966), the greatest CS intensity effect occurring with the 500-msec. ISI.

These results suggest that it would be fruitful to explore CS intensity and ISI factors in differential conditioning. In addition, it has been suggested that CS intensity effects in eyelid conditioning are related to the contrast between the intensity of the CS and the intensity of the background against which the CS is presented rather than to the absolute CS intensity (e.g., Moore, 1964). There has been no systematic investigation of this possibility, although such a finding could have theoretical implications and be of methodological importance as well since background noise levels have varied in studies investigating CS intensity.

---

[1] This investigation was supported by United States Public Health Service Research Grant MH 10235 and United States Public Health Service Training Grant 5 T01 HD00 117 from the National Institute of Child Health and Human Development.

[2] Requests for reprints should be sent to Leonard E. Ross, Department of Psychology, Psychology Building, Charter at Johnson, University of Wisconsin, Madison, Wisconsin 53706.

The present study was designed to investigate the between-*S*s effects of CS intensity, background intensity, and ISI in a design in which the three variables were orthogonal. With the loudness of CS+ and CS− matched, tone and noise intensities were selected such that there would be two CS intensity increments above each of two background white noise intensity levels. With two ISI values orthogonal to background level and CS intensity increment, there were a total of four groups at each background intensity (65 and 80 db.), two for which CS+ and CS− were 5 db. above background intensity and two for which CS+ and CS− were 20 db. above background intensity. With this design, CS intensity or ISI effects could differ as a function of absolute CS intensity or as a function of CS intensity relative to background intensity. In the former case, differences would be expected in comparing groups that had the same CS intensity increments above background but different background intensity levels since such groups would differ in absolute intensity. On the other hand, similar effects in groups with the same CS intensity increments above background but different background levels, i.e., no significant background effect or interactions, would imply that absolute CS intensity and background intensity per se were unimportant in producing those effects.

The two levels of ISI were selected to include one value thought to be optimal for differential conditioning (800 msec.) and one shorter, presumably nonoptimal, value (500 msec.).

EXPERIMENT I

Method

Subjects.—Ninety-six college students enrolled in a course in introductory psychology at the University of Wisconsin participated in Exp. I in order to earn experimental "points." The *S*s were randomly assigned to the eight experimental groups, with the restriction that each group contain an equal number of males and females.

Apparatus.—The study was conducted in a two-room mobile laboratory parked on the University campus. The *S* was seated in an ophthalmologist's chair equipped with an adjustable headrest and faced a movie screen on which silent films were shown throughout the experimental session.

The sound system included two Hewlett-Packard ABR audio oscillators, each generating one of the two tone CSs, a Grason-Stadler white-noise generator, which provided the continuous background white noise, and a Grason-Stadler electronic switch permitting control of onset and offset characteristics of the tone stimuli. All tones had a "fast" rise and decay time (about 10 μsec.), and all auditory stimuli were presented binaurally to *S* through Sharpe HA-10A earphones.

One of two white-noise levels was continuously presented to each *S*. The intensity values for these levels were determined by four judges who matched the loudness of white noise in the earphones to the loudness of free-field white noise of 65 and 80 db. (re .0002 dynes/cm^2).

The tone CSs were at one of two loudness levels above white noise, and again, the intensity values for these tones were based on loudness matching by the judges who matched the given frequency in the appropriate white-noise background to a 1,000-Hz. tone at the required intensity above threshold. For example, to determine the intensity settings for the cell involving white noise at 65 db. and a CS intensity of 5 db. above white noise, the loudness of both an 800-Hz. and a 2,100-Hz. tone in a background of white noise at 65 db. was matched to a 1,000-Hz. tone at an intensity of 70 db. above threshold. The resulting intensity, then, was 70 db. SL. This procedure was followed for each background-intensity-CS-intensity combination except for the combination of 80-db. background and +20-db. CS intensity, where a value of 94 db., rather than 100 db., was selected on the basis of published equal-loudness contours (Stevens, 1951). Such a procedure was used so that the loudness differences between the +5-db. CSs and the +20-db. CSs would be equal at both background intensity levels. Judgments confirmed this value insofar as the difficult difference judgments could be made.

The UCS was a puff of compressed air delivered to the right cornea with an intensity of .75 psi and a duration of 100 msec. The CS and UCS terminated together.

Eyelid responses, both direct and differentiated signals, were recorded by the usual potentiometer-polygraph arrangement.

Intertrial intervals were 15, 20, and 25 sec., with a mean of 20 sec. All sound-generating and response-recording equipment and the movie projector were located in a second room of the trailer, where a continuous background noise level was maintained in order to mask any equipment sounds.

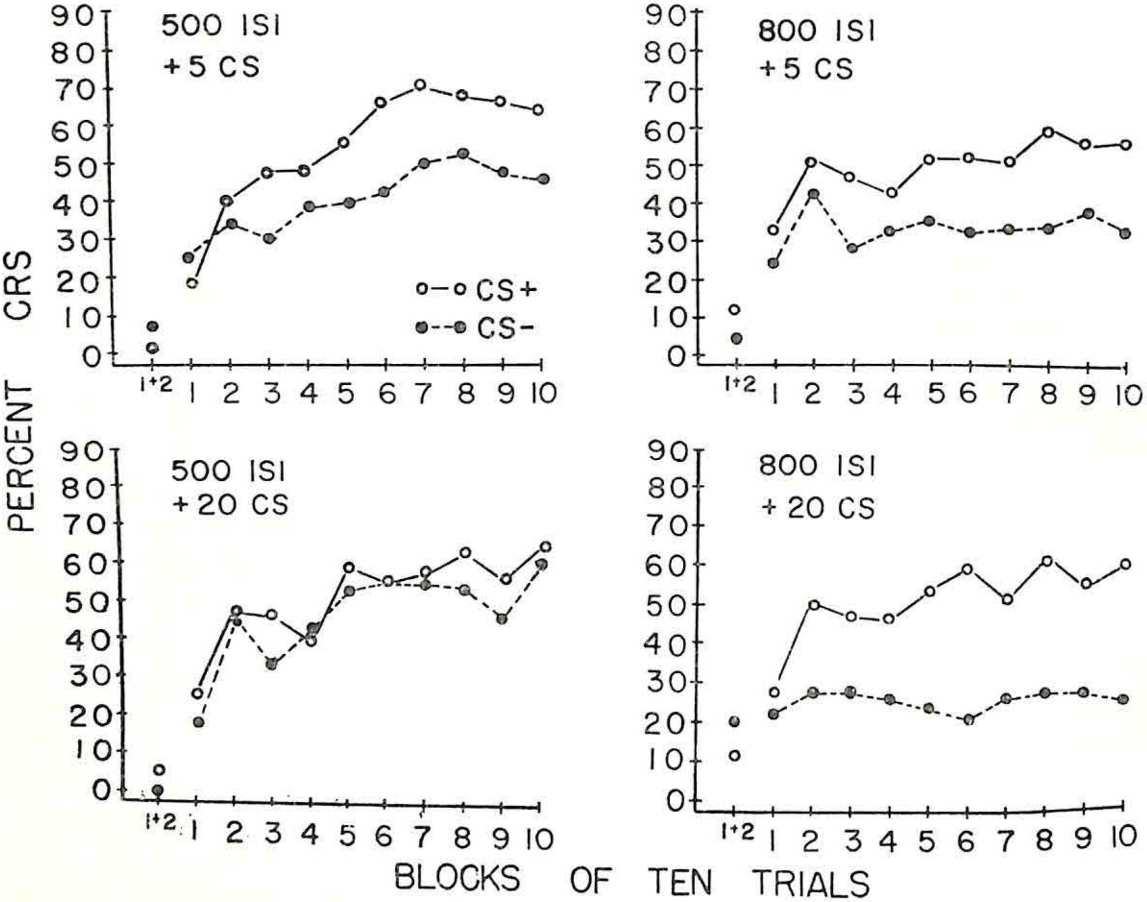


FIG. 1. Percentage of CRs in blocks of 10 trials for the four ISI–CS-intensity combinations summed across background intensity levels. (Each group contains 24 *S*s.)

Design and procedure.—The design of the study was a 2×2×2 factorial. Conditions included two levels of ISI (500 and 800 msec.), two loudness levels of white noise (65 and 80 db.), and two relative CS intensity levels (5 and 20 db. above white noise). The two tone CSs, 800 and 2,100 Hz., were presented in a Gellermann order such that no more than three presentations of the same tone occurred successively. For half of the *S*s, the 800-Hz. tone was reinforced (CS+) and the 2,100-Hz. tone was not reinforced (CS−). For the remaining half of the *S*s, these contingencies were reversed.

After *S* had been seated in the experimental room, he was informed that the equipment was to be used to measure his reactions to certain events as these events occurred while he was watching a silent motion picture. He was to relax, watch the movie, and make no attempt to aid or inhibit his natural responses. The instructions made no mention of the puffs or tones to be presented. The headset and earphones were then fitted, and *E* left the experimental room and remained in the equipment room for the duration of the conditioning session, during which each *S* was given 100 trials of differential eyelid conditioning.

Results

The first pen deflection of 2 mm. or more (.67 mm. at the eye) occurring within the interval beginning 200 msec. after CS onset and ending with the onset of the UCS was scored as a CR. Although this procedure results in a 300-msec. scoring interval for the 500 ISI group and a 600-msec. scoring interval for the 800 ISI group, no correction for scoring interval length was made because difference scores were analyzed and responding to both positive and negative cues might be expected to be similarly affected by random blink rates.

An analysis of variance of these data was performed using difference scores summed over the last four blocks of trials and including as sources of variance sex of *S*, frequency used as CS+, white-noise intensity level, ISI, and CS intensity. There was no significant effect of sex of *S* ($F < 1$), frequency used as CS+ ($F < 1$), or any of the interactions involving these factors. In addition, the data for the two background white-noise intensities were quite similar, with neither the effect of white-noise background intensity ($F < 1$) nor any of the interactions involving this factor significant. Figure 1 shows the percentage of CRs in blocks of 10 trials for the four ISI–CS-intensity combinations that resulted when groups were combined over all nonsignificant factors. Each group contains 24 *S*s.

The ISI × CS Intensity interaction, apparent in Fig. 1, was confirmed by the analysis, $F\ (1, 64) = 4.51,\ p < .05$. Subsequent *t* tests indicated better discrimination by the 800 ISI group at the +20 CS intensity level than by the 500 ISI group at the same CS intensity level, $t\ (64) = 3.14,\ p < .01$. There was no ISI difference at the +5 CS intensity level, $t\ (64) = .13,\ p > .10$. Neither the CS intensity difference at the 500 or the 800 ISI value was significant, $t\ (64) = 1.64,\ p > .10$; $t\ (64) = 1.36,\ p > .10$, respectively, although the difference at the 500 ISI level approached the .10 level.

In order to distinguish voluntary form responders (Vs) from nonvoluntary form responders (Cs), conditioned responses were scored according to the time-derivative criterion suggested by Hartman and Ross (1961). Two separate unweighted means analyses of variance, including ISI and CS intensity as sources of variance, were carried out over the summed difference scores for the last four trial blocks to determine whether the ISI × CS Intensity interaction previously observed was specific to either the V or C groups. Resulting *F* tests found a significant CS Intensity × ISI interaction for Cs, $F\ (1, 66) = 5.85,\ p < .025$, but only a significant ISI main effect for Vs, $F\ (1, 22) = 24.38,\ p < .001$.

Discussion

In Exp. I there was no evidence that the absolute intensity of the white-noise background affected the final level of differential performance, and the absence of Background Level × CS Intensity interaction leads to the conclusion that the CS intensity interaction was a function of CS intensity relative to background level rather than to absolute CS intensity since tones heard above 80-db. white noise were of a greater absolute intensity than tones heard above 65-db. white noise.

There was no difference in differential conditioning between the two ISI groups at the +5 CS intensity level, but significantly better differential performance by the 800 ISI group as compared to the 500 ISI group at the +20 CS intensity level. There was also a marginally significant difference, for Cs alone, between the two CS intensity groups at the 500 ISI level. In this latter case, the effect is in the opposite direction to what might be predicted since the group differentially conditioned to stimuli of greater intensity relative to the background showed poorer differential performance than the group differentially conditioned to stimuli less intense relative to the background.

These results suggest an interpretation based on the onset characteristics of the CS pair. Since the tones had a "fast" rise time and since the earphones would not be expected to follow such an onset, transients heard as a "click" would be expected at the onset of the tone. These transients would be similar for both CS+ and CS− and could tend to increase the similarity of the two tones, especially if the tone onset and its transient artifact characteristics were the major effective conditioning cue. If it can be assumed that these transients are masked by the white-noise background and if the degree of masking is greater when the tone and the white noise are close together in intensity, then it would follow that the two cues were more alike for the +20 CS intensity groups than for the +5 CS intensity groups. Under these circumstances, discrimination on the basis of frequency would be more difficult at the +20 CS intensity level.

Assuming that there is some time requirement for the inhibitory or perceptual processes that are presumably necessary for differential responding, the significant difference between the two ISI groups at the +20 CS intensity level follows. Although the masking effect of the white noise and the onset characteristics of CS+ and CS− were the same for the two groups, the lengthened ISI may have permitted the necessary discrimination processes and therefore differential responding to the two tones to occur.

At the 500 ISI level, the difference between the two CS intensity groups indicates that as relative CS intensity is increased, differential performance becomes poorer, presumably due to a decrease in the masking of onset transients. When the same comparison is made at the 800 ISI level, however, the difference is no longer apparent. Again, it appears that at 800 ISI, the time interval is sufficient to allow the +20 CS intensity group to differentially respond before the end of the scoring interval.

While the interpretation in terms of the masking of onset transients is admittedly speculative, it appears to fit the data well and is plausible in view of Wickens and Wickens (1942) demonstration that stimulus rise time can be an effective conditioning cue. Manipulation of the transients by variations of the rise time of tones presented against a background of white noise should lead to results similar to those of this study, particularly when the shorter ISI is used.

EXPERIMENT II

In order to investigate the relationship between CS onset characteristics and ISI suggested by the results of Exp. I, a second experiment was designed using two levels of ISI and two values of rise–decay time for the tone CSs. Having assumed that the onset characteristics of the CS are a critical factor in differential responding, it was expected that reducing the onset transients by means of increasing the CS rise–decay time would lead to improved performance at the short ISI (500-msec.). No appreciable effect would be expected at the longer ISI (800-msec.) since it was predicted that here the increased time between stimuli would allow *S* to compensate for stimulus onset similarities and to differentially respond before the end of the scoring interval. Only +20 CSs were used in Exp. II because at the +5 CS intensity level, onset transients were assumed to be masked already by the background white noise to such an extent that the onsets of CS+ and CS− were not highly similar.

Method

Subjects.—Forty-eight college students enrolled in introductory psychology participated in Exp. II in order to earn experimental "points." The *S*s were randomly assigned to the four experimental groups, with the restriction that each group contained an equal number of males and females.

Apparatus.—The study was conducted in a three-room laboratory suite, which allowed two *S*s to be tested simultaneously. Each *S* was seated in an ophthalmologist's chair in a separate room and faced a movie screen on which silent films were shown throughout the experimental session.

The sound system included the same types of components described for Exp. I and was calibrated in exactly the same manner by a different group of four judges. The procedure was similar to that used in Exp. I, with the exception that the mean intertrial interval was 15 sec. The UCS was of the same intensity and duration used in Exp. I, and the CS and UCS terminated together.

Response recording equipment was the same as that used in Exp. I and, along with the sound-generating apparatus, was located in a separate room.

Design and procedure.—The design of Exp. II was 2 × 2 factorial, with conditions including two levels of ISI (500 and 800 msec.) and two values of rise time ("fast" and 50 msec.). The two tone CSs were 800 and 2,100 Hz. and were presented with an intensity of +20 db. against background white noise at 65 db. The order of CS+ and CS− trials was the same as that used in Exp. I, and the same counterbalancing procedure for the frequency of CS+ was used. After receiving the instructions previously described, each *S* was given 100 trials of differential eyelid conditioning.

Results and Discussion

Records were scored according to the same criteria used in Exp. I, and Fig. 2 shows the percentage of CRs in blocks of 10 trials for the four ISI–rise-time combinations. Each group contained 12 *S*s. Differences predicted at the outset of the experiment were tested by means of *t* tests. As expected, the 800-ISI–fast-rise-time group performed better than the 500-ISI–fast-rise-time group, $t\ (22) = 2.52$, $p < .01$, replicating the results of Exp. I. There was no difference between the two rise-time groups at the 800 ISI level, $t\ (22) = .28$, $p > .10$, but at the 500 ISI level, as predicted, the slow-rise-time group performed better than the fast-rise-time group, $t\ (22) = 2.28$, $p < .025$.

The time-derivative criterion was used to separate Vs from Cs, with the distribution of Vs as follows: 500-slow, 9; 500-fast, 6; 800-slow, 4; 800-fast, 4. However, results for these two groups were not analyzed

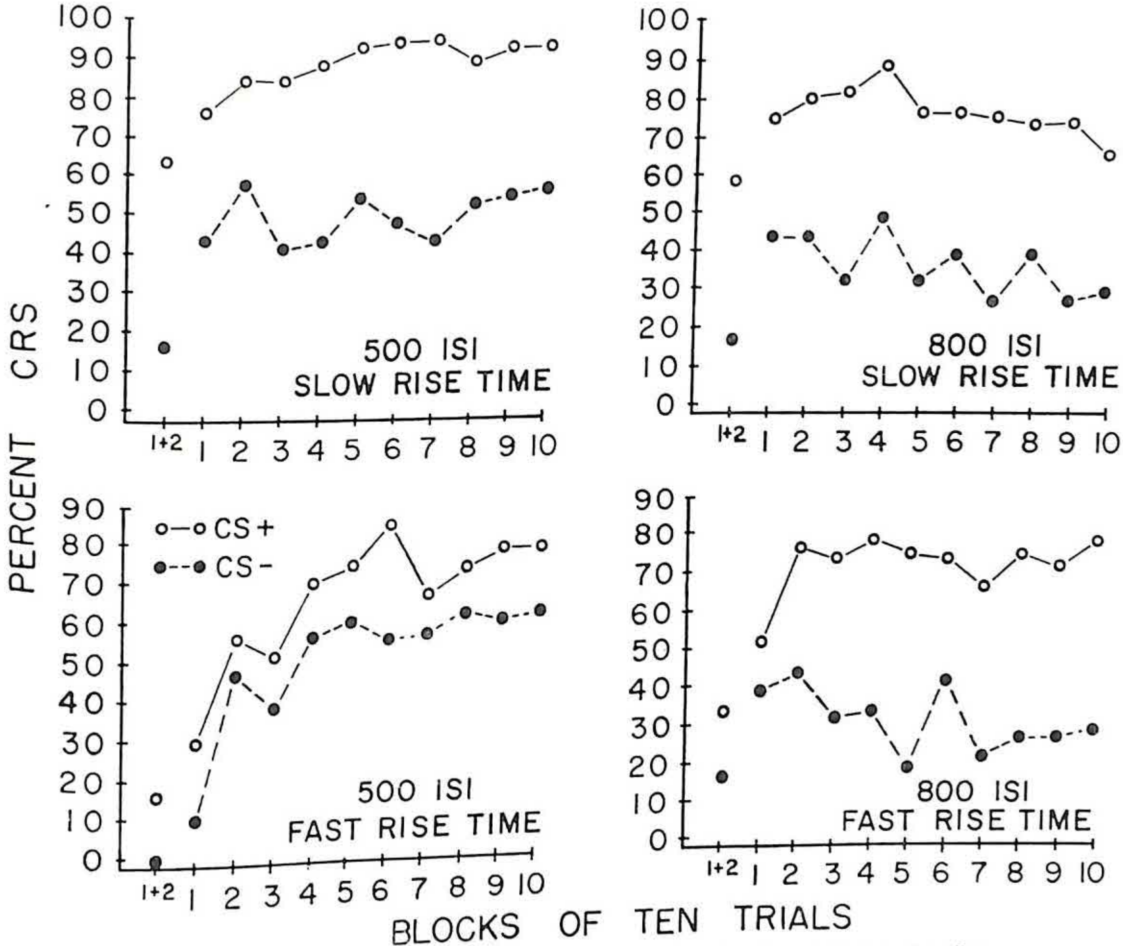


FIG. 2. Percentage of CRs in blocks of 10 trials for the four ISI–CS-rise-time combinations. (Each group contains 12 *S*s.)

separately because of an unusually large number of Vs in the 500-ISI–slow-rise-time group. Two more groups of 12 *S*s each were tested at the 500 ISI under slow and fast rise time in order to obtain a larger number of Cs in those two conditions. With the additional *S*s, there were 11 Cs in the fast-rise-time group and 10 Cs in the slow-rise-time group. A t test indicated that for Cs alone, slow rise time led to more differential responding than the fast rise time, t (19) = 2.30, $p < .025$. For the expanded V groups, there was no difference between the two levels of rise time, t (21) = .69, $p > .10$.

The finding that an increase in rise–decay time of auditory CSs leads to improved differential performance at the short ISI but not at the long ISI is entirely consistent with the expectations based on the outcome of Exp. I. Presumably, increasing the rise time of tone CSs reduces the similarity between CS+ and CS− onsets by reducing onset transients in cases where CSs are sufficiently louder than the background for transients to be present when rise time is uncontrolled. It is obvious, in Fig. 2, that this manipulation brings the 500-ISI–slow-rise-time group to the same level of differential performance found in either of the two 800 ISI groups, suggesting that there is a trade-off between CS onset characteristics and ISI.

It is possible that the intensity, rise time, and ISI effects observed in the present studies are related to the orienting response (OR) processes. Relationships between the OR and conditioning have been postulated (Sokolov, 1963), and OR components have been found to be related to tone intensity and the onset characteristics of pure tone stimuli (Graham & Jackson, in press). However, the nature of the relationship between the OR and conditioning has not been specified to such an extent that the results of the present studies can be interpreted within such a framework.

REFERENCES

CARTER, L. F. Intensity of conditioned stimulus and rate of conditioning. *Journal of Experimental Psychology*, 1941, 28, 481–490.

GRAHAM, F. K., & JACKSON, J. C. Arousal systems and infant heart rate responses. In L. P. Lipsitt & H. C. Reese (Eds.), *Advances in child development*. Vol. 5. New York: Academic Press, in press.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and the strength of conditioning: I. The conditioned eyelid response to light. *Journal of Experimental Psychology*, 1948, 38, 690–696.

HARTMAN, T. F., & ROSS, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *Journal of Experimental Psychology*, 1961, 61, 334–338.

LIPKIN, S. G., & MOORE, J. W. Eyelid trace conditioning, CS intensity, CS-UCS interval, and a correction for "spontaneous" blinking. *Journal of Experimental Psychology*, 1966, 72, 216–220.

MATTSON, M., & MOORE, J. W. Intertrial responding and CS intensity in classical eyelid conditioning. *Journal of Experimental Psychology*, 1964, 68, 396–401.

MOORE, J. W. Differential eyelid conditioning as a function of the frequency and intensity of auditory CSs. *Journal of Experimental Psychology*, 1964, 68, 250–259.

SOLOLOV, YE. N. *Perception and the conditioned reflex*. New York: Macmillan, 1963.

STEVENS, S. S. (Ed.) *Handbook of experimental psychology*. New York: Wiley, 1951.

WALKER, E. G. Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *Journal of Experimental Psychology*, 1960, 59, 303–311.

WICKENS, D. D., & WICKENS, C. D. Some factors related to pseudoconditioning. *Journal of Experimental Psychology*, 1942, 3, 518–526.

(Received February 25, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 279–284

RELATIONS BETWEEN THE SENSORY REGISTER AND SHORT-TERM STORAGE IN TACHISTOSCOPIC RECOGNITION[1]

A. O. DICK[2]

University of Waterloo

A partial report procedure was used in which groups of *S*s were required to report elements of a set of stimuli according to one of three attributes: spatial location, color, or class (letters and numbers). The results indicate differential loss of information as a function of delay of report, presumably a failure to transfer some of the information from the sensory register to short-term storage. The benefit of partial report over full report was not affected by delay, and consequently this effect seems to be the result of the short-term storage system and not of the sensory register.

Upon presentation, a stimulus is encoded in a sensory register (SR) where it exists for a brief interval before decaying. The contents of SR are recoded into different forms and some of the contents are put into short-term storage (STS) and eventually into long-term storage; the information that is not recoded into STS is apparently lost. Maintenance of information in STS is in part a function of control processes such as strategies and the like, while loss of information from SR is a function of structural properties of the system (Atkinson & Shiffrin, 1968). The experiment reported here examines some of the structural aspects of SR and some control aspects of STS.

In one of the original perceptual experiments on sensory storage, Sperling (1960) presented alphabetic and alpha-numeric materials tachistoscopically. The items were arranged in a matrix such that *E* could request report of any one of several rows. The report cue was an auditory signal presented postexposurally. The results indicate that *S* was able to report proportionally more items when asked for part of the array than when asked for the entire array and that accuracy decreased with increasing delay of the auditory signal and reached asymptote around 200 msec. This asymptote suggests that a processing mechanism operates within a finite time period (see also Estes & Wessel, 1966) and because of this time limit, information held in SR but not processed is lost, presumably as a result of neural decay.

A series of studies by Dick (1967), however, suggest that not all types of information are subject to decay; specifically, a distinction should be made between spatial and identity information. In two of these experiments, *S*s were shown two rows of letters. A partial report procedure was used in the first experiment, in which *S*s were asked to report the top row, the bottom row, or both rows on a given trial. A coded auditory cue indicated to *S* which way to report the material; this auditory cue was presented at various intervals after the exposure. The results indicate a significant decrease in accuracy with increasing delay of the cue. In a second experiment, *S*s were shown the same materials, but always were asked to report both rows. An auditory cue was used to indicate to *S* when he could begin to report the material. In other words, in the first experiment the auditory cue indicated *what row* and *when* to report, but in the second experiment the cue indicated only *when* to report. In contrast to the first experiment, there was no effect of delaying report for the letter material. The difference in the results can be explained in terms of the information that is extracted. For the first experiment, spatial information is important because *S* had to report according to spatial location (row), while in the second experiment spatial information is not crucial because *S* always had to report the entire display. Thus, the data suggest that one should make a distinction between types of information to be processed.

---

[1] This research was supported in part by Defence Research Board of Canada, Grant No. 9401-26 to M. P. Bryden. The data reported here constituted a portion of the data comprising a PhD thesis submitted to the University of Waterloo. In addition, some of these data were reported at the meetings of the Eastern Psychological Association, Boston, April 1967. The author would like to thank M. P. Bryden and D. J. K. Mewhort for critical comments.

[2] Requests for reprints should be sent to A. O. Dick, who is now at University of Rochester, Department of Psychology, Rochester, New York 14627.

The experiment reported here uses a partial report technique to assess differences in processing for several types of information. This was accomplished by arranging the stimulus materials such that several groups of *S*s could be asked to report using different attributes of the same stimulus array; however, over a series of trials, all groups reported the same items.

METHOD

Subjects.—The *S*s were 72 undergraduate arts students taking an introductory course in psychology at the University of Waterloo; participation fulfilled a course requirement. Colorblind *S*s were excluded from the experiment.

Stimulus materials.—Eight letters (A D E L M R S T) and eight digits (2–9) were used. Each stimulus card contained four letters and four numbers arranged in two rows of four, with four of the elements colored red and four of the elements colored black. The stimuli were made by placing Letraset Instant Lettering (Style 287) on white cards. The size of each item was $\frac{1}{4}$ in. high and approximately $\frac{1}{8}$ in. wide, with the overall angle subtended by the display being 4.0° horizontally × 2.0° vertically. The possible combinations of items by color by position ($\binom{8}{4} \cdot \binom{8}{4} \cdot 2 \cdot 8$) is large and obviously it was impractical to use all combinations. Instead, four sets of 18 stimulus cards were systematically prepared to incorporate examples of all combinations of identity by color by position. In all four sets, an item appeared in each of the eight possible positions at least once and in both colors approximately equally often.

Each of the four sets was randomly divided into three groups of 6 cards; the first third of each set was combined to make Block 1, the second third into Block 2, and the last third into Block 3. Within each of the three blocks the cards were randomized. In addition, 12 stimulus cards, 3 examples of each of the four sets of materials, were prepared for practice trials.

Presentation of the materials.—The stimulus materials were presented in a Gerbrands mirror-type tachistoscope. Field 1 of the tachistoscope was used to present a small black fixation point. This fixation point appeared centered between the rows of the stimulus materials, and midway between the second and third items. The stimulus cards were presented in Field 2. The luminance of Field 1 was about 3.8 ftL. and that of Field 2 was about 3.4 ftL. Field 1 was illuminated at all times except when Field 2 was exposed. In other words, a light-light-light cycle was used; use of such a cycle was intended to reduce the possibility of after-images. The exposure duration was 50 msec. (Preliminary research had shown this duration to yield accuracy of about 50% for eight-item displays.) A verbal "ready" indicated the start of a trial.

In addition to controlling the duration of exposure, the tachistoscope timers controlled an audio oscillator which produced three tones of 1-sec. duration which were used to instruct *S* how to report the stimulus material. A high tone (1,000 cps) indicated that *S* should report half of the items, a low tone (200 cps) indicated that *S* should report the other half of the array, while an intermediate tone (400 cps) indicated report of the entire display. The interval between the offset of the visual array and the onset of the tone was varied from −850 to +850 msec. Eight different intervals were used: −850, −450, −50, 50, 250, 450, 650, and 850 msec.

Within the experiment, each *S* received several randomized blocks of the delay-instruction conditions. The order in which the cards were presented was always the same; to minimize the effects of practice, the order of the delay-instruction conditions was randomized within each block and then counterbalanced across *S*s in that group. Furthermore, the settings of the timer and the oscillator were changed on every trial, even when the next trial involved the same settings.

Procedure.—The *S*s were randomly divided into three groups. Group 1 verbally reported the material by spatial location: the high tone indicated report of the top row, the low tone indicated report of the bottom row, and the intermediate tone indicated report of the entire display. Group 2 verbally reported the material on the basis of color: the high tone indicated report of red items, the low tone indicated report of the black items, and the intermediate tone indicated report of the entire display. Group 3 verbally reported the material by class of stimuli: the high tone indicated report of letters, the low tone indicated report of numbers, and the intermediate tone indicated report of the entire display. Each mode of report condition was used at each of the eight delay intervals, giving a total of 24 delay-mode combinations for each group. Each *S* was given 72 trials, three randomized blocks (replications) of the 24 delay-instruction combinations. In addition, each *S* was given 12 free recall practice trials prior to the experiment.

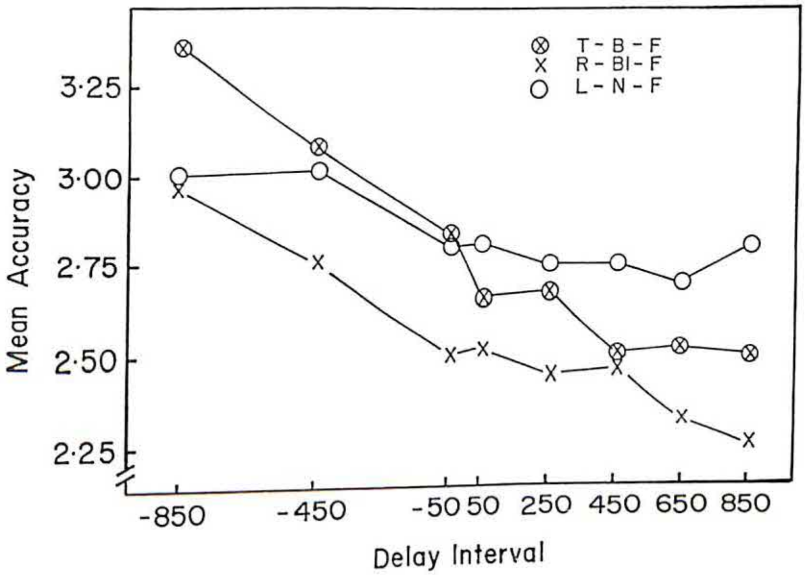


FIG. 1. Overall mean accuracy for each of three reporting conditions. (In the legend, T-B-F represents the combined average of top row, bottom row, and full report for the spatial condition; R-B1-F for the color condition; and L-N-F for class-of-item report.)

RESULTS

Overall analysis.—Using the number of items correctly identified on each trial (without regard for position), the data from the main experiment were entered into an analysis of variance (Groups × Mode of Report × Replications × Delay); this analysis indicated that the groups did not differ significantly ($p = .203$). The main effects of mode of report, replications, and delay all produced significant effects ($p = .001$), as did the Mode × Groups and Delay × Groups interactions ($p < .002$). A number of additional analyses were carried out to evaluate these significant effects in more detail. At the same time, effects can be sorted out due to different features of the system. The primary evidence related to structure is found in the postexposure delay data, while most of the other effects are related to strategies and control.

Postexposure delay.—The effect of delay was significant in the overall analysis and is shown in Fig. 1. Loss of information from SR can only occur, however, after the stimulus has been presented. Therefore, in the analysis of delay the present report deals only with the five postexposure intervals where it was possible to observe loss of information from SR. To test this, the postexposure data only were entered into an analysis of variance/trend analysis (Groups × Mode of Report × Replications × Delay). This analysis indicated a significant effect of delay ($p = .046$) and a significant linear trend ($p = .005$). To examine this effect in more detail, separate analyses were applied to the postexposure data for each group. On the basis of the trend components of these analyses, the differential effects of delay shown in Fig. 1 were quite clear-cut: there was no linear trend for class of item report ($p = .812$), a borderline effect for spatial report ($p = .061$), and a significant linear trend for color ($p = .012$). Thus, there was no evidence for a loss of class-of-item information, a hint of a loss of spatial information, and a more clear-cut loss of color information.

Superiority of partial report.—A portion of the significant effect of mode of report is a result of comparing accuracy on the partial report trials where only four responses were possible with accuracy in full report where eight responses were possible. Mode of Report × Groups interaction, however, indicates that other things may be occurring, and therefore a comparison of the accuracy for the report conditions should provide additional information about the mechanisms involved. That is, one can examine the superiority of partial report over full report; this is done by summing the number of correct responses for each of the two partial report conditions and comparing this sum to the accuracy in full report. For example, in the spatial report group the means at 650 msec. were 2.67 for the top row and 1.71 for the bottom row. The sum of these two conditions, 4.38, was then compared to the accuracy for full report, which in this case was 3.19. The two partial report scores for each *S* in each group on each trial were summed and then compared to full report in an analysis of variance for each group independently.

For all groups, partial report accuracy was higher at all intervals than full report accuracy ($p = .001$); however, none of the Mode of Report × Delay interactions were significant. The superiority of partial report in the class condition was of particular interest since this is apparently not consistent with the findings of other investigators (e.g., Beach & Bryden, 1963; Sperling, 1960). Neither of these previous experiments, however, employed delayed full report conditions, and hence, the present experiment provides a more precise estimate of the benefit of partial report. These results have implications for the way in which information is handled in STS and because of the magnitude of the effect, it is likely that in addition to the possibility of output interference, other effects are involved such as interference due to rehearsal. This possibility is discussed in more detail later.

Other effects.—Another comparison that can be made is between the partial report conditions within each group. Such a comparison reveals that accuracy for (*a*) the top row is reported significantly higher than the bottom ($p = .001$); (*b*) red is significantly higher than black ($p = .001$); but (*c*) letters are not significantly different from numbers ($p = .066$). However, only for the spatial condition is the interaction between the two modes of partial report and delay significant ($p = .004$). Assuming that the elements are processed from left-to-right and top-to-bottom, the spatial information for the bottom right element in the display would suffer more decay because it would be processed last, if at all. For example, if the tone occurred early, the bottom row could be processed before decay was complete; however, if the tone occurred late, decay would already have taken place. Therefore, the bottom row should show greater effects of delay than the top row; the significant interaction supports this interpretation.

DISCUSSION

In summary, groups of *S*s were asked to report by spatial location, color, or class of material, using a partial report procedure; three main findings were observed. (*a*) Decreases in accuracy with increasing delay were observed postexposurally for spatial report and for color, but not for class report. (*b*) Accuracy was always higher for partial report than for full report. (*c*) There is a bias observed in all groups favoring the top row, red items, and to some extent letters, as well as greater accuracy on preexposure cued trials.

Atkinson & Shiffrin (1968) have proposed a three-stage model of memory which can be employed to account for the present findings. In this model the three stages are (*a*) a sensory register (SR) in which information is represented in photographic-like form, (*b*) a short-term storage (STS), and (*c*) a long-term storage. Information decays from SR and STS; however, the form of representation differs for the two. A visual stimulus in SR might be represented in a visual code, while in STS the same stimulus might be represented in an auditory code. Although little is known about the structural properties of the first two stages and the interrelations between stages, the present experiment together with other data can be used to provide some hints about these mechanisms. Following Atkinson and Shiffrin, structural properties are first discussed and then control processes.

Structural properties.—One of the problems in understanding the relation between stages of perceptual processing is the lack of consistency in results of experiments employing different kinds of stimuli. An experiment pointed at the resolution of some of the inconsistency shows that more familiar 2-order approximations to English are transferred from SR (iconic memory) to STS faster than are less familiar 0-order approximations (Mewhort, Merikle, & Bryden, 1969). While the specific mechanisms are not fully understood, these data are clear in showing that the rate of processing varies as a function of the stimulus input. If one is willing to assume a parallel between sequential information and structural information, then the postexposure delay differences in the present experiment between the color, spatial, and class-of-item report groups can be easily understood.

In the present experiment, all *S*s were required to report identity information in free recall. It would seem reasonable that this type of information should be transferred at the same rate for all groups. However, the groups differ in the type of additional analysis that must be carried out. The class-of-item group must first identify or name the item before classifying it. As such, class of item is derived information, not stimulus information. This interpretation is supported by the reaction time data of Posner & Mitchell (1967). In their experiments they showed that reaction time for same–different judgments is longer when *S* is asked to make the judgment on the basis of vowel-consonant than when made on the basis of names of the items.

By contrast, in the other two groups, spatial and color report, *S*s were asked to respond on the basis of physical aspects of the stimulus. The additional requirements of physical information should require more processing per element and therefore slow the rate of processing.[3] A slower rate of processing would be reflected in a lower level of accuracy—an effect that was found, although these differences were by no means statistically significant. More importantly, however, one would expect differential effects as a function of postexposure delay at the report cue, an effect that was statistically significant. This latter effect can be explained in the following fashion.

First, the added requirement of physical information (color and spatial) would require additional processing or transfer time per element. Second, both physiological estimates (e.g., Kooi & Bagchi, 1964) and behavioral estimates (e.g., Sperling, 1960) suggest that the total processing time available is about 250 msec. Based on the Mewhort et al. (1969) data, processing time per letter is about 20–30 msec. (for 0-order) and thus all of the identity information could barely be transferred within this interval. In fact, for the class report group, identity would be the only information that would be necessary from SR. While it would be necessary for *S* to classify the item after he identified it, this could be accomplished in STS after transfer had occurred. Since it is possible that *S* could transfer all of the items on the basis of identity before complete decay occurred, no decrease in accuracy should be observed. When either spatial or color information is added to this requirement, processing time per element must necessarily be longer, and therefore the total time needed would exceed the period available; hence one would obtain a decrease in accuracy. In other words, the more information that must be processed per element, the longer the time required to do the processing, and therefore the time required to process a constant number of items would increase with a consequent decrease in accuracy under the present conditions.

It is relatively clear that *S* will process elements in a left-to-right, top-to-bottom manner in the absence of other instructions (e.g., Dick, 1967; Mewhort et al., 1969). Furthermore *S* does not appear to wait for the report cue before he begins to process, otherwise accuracy would be zero for long delays; and therefore, under partial report conditions, sometimes *S* will process the "wrong" elements. Consider, e.g., the case in which the cue indicates report of the bottom row and occurs at 450 msec. after the exposure. The *S* would have processed identity and location for most of the elements from the top row and for a few elements from the bottom row. Because decay would be complete by the time the tone occurred, *S* would be unable to carry out further processing and would have to report whatever he had been able to process from the bottom row. This is reflected in the main effect of mode of report and the Groups × Mode of Report interaction. The difference between partial report cues within each group of *S*s seems to be due to a bias on scanning in transfer. By contrast, the differential accuracy between groups seems to be due to a finite time limit on processing and differential rates of processing.

---

[3] While this argument may not appear to hold for the prior cue conditions, other factors also have to be considered. For example, in the spatial condition with the cue 850 msec. prior to the exposure, *S* has ample time to move his eyes to the appropriate row. This explanation is partially supported by the interaction for the spatial condition of the modes of partial report by delay. This is not, however, a complete explanation of the pre- and postcuing differences (e.g., Haber, 1966).

Control processes.—According to Atkinson and Shiffrin (1968), control processes include strategies such as set or attention and rehearsal which they define simply as anything that postpones decay of information from STS. Several of the findings in the present experiment are relevant to a discussion of control processes. For example, one apparent simple strategy of each *S* in the present experiment is that he will look at the top row and report that first, and he apparently attends to the red items. Such an effect seems to be due to both familiarity and specific experience with the material (Bryden, Dick, & Mewhort, 1968).

The benefit of partial report can be interpreted in terms of control processes operating in STS for three main reasons. First, the benefit of partial over full report was observed in the letter-number report condition even though no loss of information was observed. Second, while one might want to attribute the benefit to differences in output interference, the data of Dick and Mewhort (1967) provide no consistent evidence for the notion of output interference. Third, the benefit occurring after asymptote cannot be attributed to SR because information in SR should have decayed completely. That is, if SR were responsible for the superiority of partial report, one would expect either an interaction of accuracy in partial and full report with delay or one would expect a nonlinear effect because of transfer from SR to STS. Clearly, there is no evidence that the benefit of partial report is a function of timing of the report cue either prior to or after the exposure; therefore it is consistent to suggest that a mechanism is involved that is different from the one responsible for the decrease in accuracy with postexposure delay. Following from Atkinson and Shiffrin (1968), such a mechanism might be rehearsal in STS.

Rehearsal could influence the benefit in the following way: By specifying the items to be reported, *S* no longer has need of the remaining items, need not rehearse them, and therefore can allow them to decay from STS. Since these items no longer need to be rehearsed, more rehearsal time can be devoted to maintaining the items that are needed. This aspect of STS could influence partial report accuracy through a reduction in memory load, i.e., with fewer items to keep track of, loss of information from STS would be considerably reduced and hence accuracy should be proportionally higher. This interpretation does not necessarily deny the possibility of rehearsal in SR or during transfer. However, it is not clear how such rehearsal could account for the differential effects of delaying the report cue.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

Bryden, M. P., Dick, A. O., & Mewhort, D. J. K. Tachistoscopic recognition of number sequences. *Canadian Journal of Psychology,* 1968, 22, 52–59.

Beach, B. L., & Bryden, M. P. The effects of pre- and post-exposure instructions on tachistoscopic recognition. Paper presented at meetings of the Canadian Psychological Association, Quebec City, June 1963.

Dick, A. O. *Visual information processing and the memory trace.* Unpublished doctoral dissertation, University of Waterloo, 1967.

Dick, A. O., & Mewhort, D. J. K. Order of report and processing in tachistoscopic recognition. *Perception and Psychophysics,* 1967, **2**, 573–576.

Estes, W. K., & Wessel, D. L. Reaction time in relation to display size and correctness of response in forced-choice visual signal detection. *Perception and Psychophysics,* 1966, 1, 369–373.

Haber, R. N. Nature of the effect of set on perception. *Psychological Review,* 1966, **73**, 335–350.

Kooi, K. A., & Bagchi, B. K. Visual evoked responses in man: Normative data. *New York Academy of Science, Annals,* 1964, **112** (Art. 1), 254–269.

Mewhort, D. J. K., Merikle, P. M., & Bryden, M. P. On the transfer from iconic to short-term memory. *Journal of Experimental Psychology,* 1969, **81**, 89–94.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. *Psychological Review,* 1967, **74**, 392–409.

Sperling, G. The information available in brief visual presentations. *Psychological Monographs,* 1960, **74**(11, Whole No. 498).

(Received February 27, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 285–289

ARBITRARY CVC HIERARCHIES LEARNED BY PAIRED-COMPONENT PRESENTATIONS[1]

EDWARD J. CROTHERS[2]

University of Colorado

The *S*s learned sets of eight statements of the form "An A is a type of B," where A and B were CVC syllables. Although the statements were presented serially, each set embodied a hierarchial organization, representable as a tree graph. Recognition learning was faster when the presentation order grouped horizontal tree levels on successive trials than when the order reflected the vertical subordinate–superordinate tree paths. Also, more learning occurred when the two doubly reinforced statements on each cycle involved the most superordinate CVC than when they involved an intermediate-level CVC. However, no dependencies were found between the learning of different statements pertaining to the same horizontal level or vertical path.

Verbal hierarchies have been studied in order to determine how familiar words are organized in memory (Deese, 1965, pp. 154–156; Mandler, 1968, p. 113; Miller, 1967) and retrieved from it (Bower, Clark, Lesgold, & Winzenz, 1969). But the original learning of a hierarchy has seldom been investigated. In the present experiment, the aim was to investigate acquisition using a hierarchy whose graphic form is a tree (see Fig. 1). Each tree contained seven nodes, and each node name was a CVC. An arrow from, e.g., SUR to PIL means that *S* is to learn some hypothetical relation between SUR and PIL. This experiment involved only one relation, "is a type of."

Instead of displaying all elements of a tree concurrently as a graph, information was presented about one pair of CVC nodes at a time in order to see how learning depends on the serial order and frequency of presentation. These two variables can be described best by defining "subpaths," "paths," and "levels" of the Fig. 1 tree. Each pair of nodes joined by an arrow forms a (directed) subpath. A single serial presentation corresponded to one subpath. For example, on a reinforced trial for SUR → PIL, *S* saw the statement "A SUR is type of PIL." (Each statement may be regarded as an S-R pairing, SUR-PIL in thi case.) A path is a unidirectional sequenc of two subpaths, e.g., from SUR to CUF bu not from SUR to FAM. A level is any one o the three horizontal rows of the graph. In brief, the presentation orders corresponded to different ways a person might "scan" Fig. 1. One way is to proceed path by path, going left to right between paths and, e.g., upward within each path (e.g., SUR → PIL, PIL → CUF, FAM → PIL, PIL → CUF again, . . . , HAV → CUF). Another way is to go level by level, moving left to right within levels and either upward (SUR → PIL, FAM → PIL, BOL → HAV, . . . , HAV → CUF) or downward (PIL → CUF, HAV → CUF, . . . , DEP → HAV) between levels. Another pos-

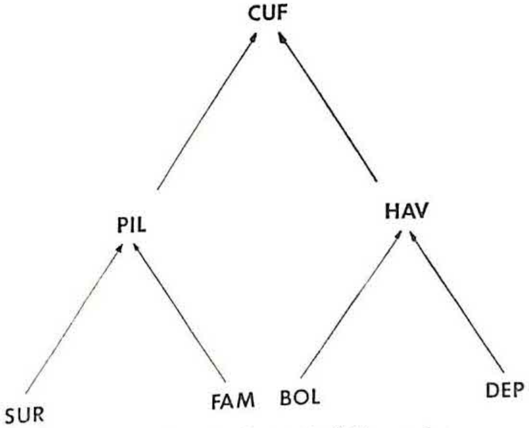


Fig. 1. Graph of a typical hierarchy.

---

[1] This study was partially supported by a grant from the Council on Research and Creative Work of the University of Colorado and is reported as Publication Number 7 of the Institute for the Study of Intellectual Behavior.

[2] Requests for reprints should be sent to Edward J. Crothers, Department of Psychology, University of Colorado, Boulder, Colorado, 80302.

TABLE 1

Presentation Frequency and Serial Order on a Reinforced Cycle

Subpath	Group			
	LU	PU	LD	SLD
Lower				
Left				
SUR → PIL	1	1	5	3,5
FAM → PIL	2	3	6	4,6
BOL → HAV	3	5	7	7
Right				
DEP → HAV	4	7	8	8
Upper				
Left				
PIL → CUF	5,7	2,4	1,3	1
Right				
HAV → CUF	6,8	6,8	2,4	2

Note.—Numeral(s) indicate the trial number(s) within each eight-trial cycle. For Group R, a new random order was used with each cycle and *S*.

sibility is a random order. An inherent property of the "paths" order, and hence a necessary control in the other orders, is that both higher level subpaths appeared twice and each of the four lower level subpaths once. To examine presentation frequency, another "levels" condition was added in which the repeated subpaths converged on a subordinate node (SUR → PIL and FAM → PIL) rather than on the superordinate node CUF.

Test cycles used a recognition method, modified to permit concurrent testing of all items so as to minimize effects of the testing order itself (Norman & Waugh, 1968).

METHOD

Design.—There were five groups of *S*s, each serving under a combination of one order condition with one frequency condition, but with three reinforcement test cycles per CVC tree, and with replications on two more trees having different CVC node names. For four of the groups, the two extra presentations on each reinforced cycle involved the superordinate CVC; e.g., one was PIL → CUF and the other was HAV → CUF. These groups are designated according to their presentation order as follows: levels up, LU; levels down, LD; paths up, PU; and random, R. For the fifth group, the two extra reinforcements per cycle were assigned to a subtree, e.g., SUR → PIL and FAM → PIL. The LD order was used, so this group is called "subtree, levels down" (SLD). Except for Group R, each group's presentation conditions were constant over cycles and trees. Hence the Table 1 illustration suffices to specify these conditions. The order of receiving a particular tree (Cycles 1–3, 4–6, or 7–9) was counterbalanced over *S*s.

Materials.—The other two trees had the same graph structure as the Fig. 1 tree but different CVC syllables as node names. Criteria for selecting the seven CVC syllables to serve as node names for a particular tree were that the Glaze association value be 93–100% and that no two syllables have the same initial or final consonant. Assignment of CVC names to nodes was the same for all *S*s and was random save for two restrictions: The vowel member of the name was not the same for any two nodes joined by a subpath; and the name of the topmost node shared no letters with the name of the topmost node in the other two trees.

Each tree had six subpaths, so there were six membership statements of the form "An X is a type of Y," where X and Y are CVC syllables.

For each tree, the recognition test answer sheet was a 7 × 8 matrix, with row and column headings typed in, but cell entries left blank. Row headings, and likewise column headings, were the seven CVC names in a random order, which varied from one test cycle to the next but was identical for all *S*s. The eighth column was headed by the word "none" so that if *S* thought a particular CVC was not "a type of" any other CVC, he could so indicate.

Procedure.—Each *S* was run individually. The initial instructions included demonstrations with a tree having the same graph structure as the trees used in the experiment but different CVC names. He was told, e.g., that "A BEZ is a type of LOR" is a statement of class membership, just like the statement "Psychology is a type of science," except that the CVC statement describes a purely hypothetical relationship. He was instructed to remember "What is a type of what." Each reinforced cycle consisted of eight trials, six plus the two repeated subpaths, in the Table 1 order. Immediately after the eighth trial, a test was given. At the start of his first test, *S* was handed a blank answer matrix and was told to place a check mark in the (i, j)th cell if he thought that the CVC named by the Row i heading was "a type of" the CVC named by the Column j heading. Also, he was to check the $(i,$ none$)$ cell if he thought the ith CVC was not "a type of any other CVC, and he was allowed to check as many columns as he wished. Reinforced cycles alternated with test cycles until three of each had occurred per tree, then repeating with a second tree, then a third tree.

The "is a type of" statements were presented serially by means of a Kodak Carousel slide projector. Time intervals were 5-sec. exposure per reinforced trial, 1-sec. interexposure interval, and 2 min. per test cycle.

Subjects.—Ninety students from an introductory psychology class participated in the experiment as part of a course requirement. Assignment of *S*s to groups was random except that each group contained 18 *S*s, and no group was assigned its $n+1$st *S* until every group had received its *n*th *S*.

RESULTS

Several decisions had to be made about how to score the data. When "X is a type of Y" and "Y is a type of Z," *S*s might have responded with "X is a type of Z" since they were not instructed either way on this point. Based on the 162 (18 *S*s × 9 cycles) presentations of each test stimulus, the relative frequency of such responses was less than .06 for each group, except in Group PU where it was .14. It was decided to count these X-Z responses as errors because they were rarely accompanied by both X-Y and Y-Z correct responses, even in Group PU. Also, neither the response type "simultaneously checked the right answer and one or more incorrect answers" nor the type "failed to check any response to a particular test stimulus" occurred with relative frequency above .05 for any group of *S*s.

Presentation conditions.—Combining data from all nine tests, the proportions of correct responses were .79, .75, .65, .60, and .57 for the groups in this order: levels up; levels down; subtree, levels down; random; and paths up. Figure 2 plots the data separately for the first, second, and third trees that *S* learned and separately for each cycle on a given tree. Each proportion equals the number of correct responses divided by the 126 (18 *S*s × 7 nodes) presentations of test stimuli per cycle. A three-way analysis of variance was calculated for the Latin square with repeated-measures design. To produce homogeneity of variance, each *S*'s score *X* under a particular combination of the three variables was first transformed to $\sqrt{X+.5}$. The between-*S*s comparison revealed a significant effect of presentation conditions, $F\ (4,\ 75) = 13.41,\ p < .001$. The learning trend exhibited in Fig. 2 was also highly significant. However, variance associated with the particular tree was not significant, $F\ (2,\ 150) = .82$. None of the interactions approached significance, all such *F* ratios being less than one. Pairwise comparisons among groups were assessed by the Newman-Keuls method. Both Groups LU and LD were significantly superior to each of Groups SLD, R, and PU ($p < .01$ in each of the six comparisons). No other between-group differences were significant.

Interestingly, Fig. 2 learning curves for the less proficient groups do not display negative acceleration across the three cycles on a particular tree. Hence a simple linear or one-element model was rejected. Also, the figure suggests a learning-to-learn effect, even from the second tree presented to the third.

Item dependencies.—If *S*s tend to learn a path (e.g., SUR → PIL, PIL → CUF) as a unit, then the two subpaths of a path should be correlated in terms of their error frequencies. Or if instead they tend to learn a subtree (e.g., SUR → PIL, FAM → PIL) as a unit, then responses to items having a common response term should be correlated. However, 2 × 2 chi-square tests, computed separately for each tree, group of *S*s, and pair of subpaths, revealed virtually no dependencies between any pair of subpaths, even ones belonging to the same path or subtree.

This negative result does not preclude the possibility that response sequences for certain pairs of subpaths might exhibit a cycle-by-cycle correlation. For example, let C denote a correct response and E an error and consider three possible three-cycle protocols: EEC, ECC, and CCE. Assume that all three come from the same tree and *S*. Informally speaking, the first of these three resembles the second more than it does the third, even though the latter two have identical error frequencies. Ideally, what we would hope to find is that the first two protocols are data from the two subpaths of the same path or subtree, whereas the third protocol is for some remote subpath. Two analyses were performed, one looking for path learning and the other looking for subtree

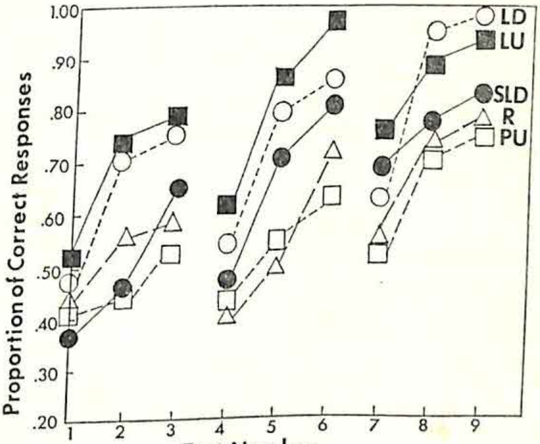


FIG. 2. Response proportions for the three test cycles on each tree.

TABLE 2

Proportions of Correct Responses at each Serial Position

Group	Position								
	0	1	2	3	4	5	6	7	8
LU	.64	.89	.73	.81	.65	.92[a]	.90[a]	.92[a]	.90[a]
LD	.68	.88[a]	.79[a]	.88[a]	.79[a]	.71	.64	.78	.77
SLD	.36	.68	.53	.75[a]	.71[a]	.75[a]	.71[a]	.76	.78
R	.41	.67	.55	.54	.54	.57	.63	.73	.76
PU	.30	.67	.73[a]	.40	.73[a]	.60	.83[a]	.48	.83[a]

Note.—Items were reinforced once, unless otherwise indicated, except that Superordinate Position 0 was not explicitly reinforced. For Group R, serial position of doubly reinforced items was random.

[a] Items reinforced twice per cycle.

learning. The details of both analyses are the same and are described for paths.

Beginning with a particular *S* and tree, the data from two subpaths belonging to the same path were compared. The number of "mismatches" was counted and then summed over the three cycles (e.g., for EEC and ECC the sum is $0+1+0=1$). But the theoretical distribution of these frequencies under the null hypothesis of independent items is unknown. Even if it were known, the finding of a significant departure from it would not be particularly interesting; the same result might be found for *any* pair of subpaths whatever. Thus, as a statistical control, data for a third subpath are required. It must be from the same *S* and tree, but remote from the other two subpaths in the graph. Moreover, it must be equated with the second (or first) subpath on the basis of the number of errors in the three cycles. Further, all protocols of the form CCC or EEE must be excluded because the number of mismatches between either of these and any other protocol is completely determined by the number of errors in the latter. This discarding of certain protocols creates the possibility that some *S*s might contribute more data to the analysis than other *S*s. However, for each *S* and path, it turned out that at least one set of three protocols meeting the above conditions could be found (i.e., either all three came from Cycles 1–3 or from Cycles 4–6 or 7–9). With *S*s for whom more than one such set was available, one set was selected randomly.

No evidence for any group of *S*s was found from this analysis to support the hypothesis that trial-by-trial correlations are higher for subpaths belonging to a common path (or subtree) than for paths that do not. Further analyses, including ones wherein the control subpath was kept constant in all comparisons for a particular subpath (e.g., for SUR → PIL the related subpath was FAM → PIL and the control subpath was BOL → HAV), corroborated this failure to find differential correlations.

Secondary results.—Table 2 gives the proportions of correct responses at each serial position. By inspecting this table along with Table 1, one can directly translate Table 2 entries into proportions correct per subpath, except for Group R where the proportions were .64, .56, .57, .50, .75, and .77 for subpaths in the order listed in Table 1. (Detailed analyses revealed no reason for the disparity between, e.g., the .64 and the .56.) Of course, only in Group R is the effect of subpath separable from the effect of serial position. The locus of the superiority of Group LD over Group SLD was early serial positions and upper subpaths. Groups R and PU were inferior to Group LU on every item and at every position except one.

Confusion matrices showed, on the whole, that errors by the more proficient groups tended to be from the same level as the correct response, whereas the poorer groups often gave responses that were not even on the right level. As noted, Group PU led in giving the superordinate node in response to bottom-level stimuli. Averaging over the three poorer groups, middle-level responses occurred to the superordinate node on 21% of the opportunities.

Discussion

There were two main findings about presentation order. First, more learning occurred if adjacent trials grouped all subpaths that were on the same level (levels up or levels down) than if they grouped subpaths belonging to the same path (paths up), and this latter order was no more effective than a random order. Second, when the grouping was according to levels, it made no difference whether the sequence proceeded from subordinate to superordinate terms (levels up) or vice versa (levels down). The main finding about presentation frequency was that repeated reinforcements were more helpful at higher (levels down) than at lower (subtree, levels down) subpaths.

Especially in the more proficient groups, it was expected that the entire tree, or at least subtrees or paths, would be learned as a unit. Dependency analyses refuted this expectation. However, the further question of whether *S*, after learning a tree, could state or sketch its graph structure was not investigated. What the data do indicate is that such knowledge, if present at all, did not influence his manner of learning subsequent trees.

What possible explanations of the main results would be consistent with this finding that subpaths were learned independently of each other? The "paths up" order inherently required that some subpaths be presented twice per cycle. Doing likewise in other orders created an unavoidable confounding between the identities and serial positions of doubly reinforced items, and thus any interpretation of these data must be tentative. Two processes might be assumed, independent learning of subpaths and guessing on unlearned items. Granted, improvement on either process might have reflected sheer practice; after all, each group did display learning-to-learn. Conceivably, however, the guessing strategy might have been sensitive to serial order and/or frequency variables. Assume that all *S*s guessed randomly on unlearned items. But assume that *S*s in the better groups gradually learned to restrict their random guesses to only the correct name and the two other names that were correct responses to some stimulus, whereas *S*s in the other groups did not restrict appreciably their set of guessing alternatives. As a first approximation, this suggestion appears compatible with the confusion matrix data as well as with the group comparisons and dependency analyses. Parenthetically, this interpretation does not necessarily imply that all groups of *S*s learned at the same rate; the rate might have varied directly with the number of response alternatives already eliminated.

The conjectured interpretation would be more nearly adequate if it suggested exactly how the presentation conditions might have produced these alleged guessing tendencies. A useful notion here is that *S*s covertly organize a list to aid their learning by grouping together certain items (Battig, 1966). With words eliciting color sense impressions as implicit associative responses, a constrained presentation order is superior to a random order (Wood & Underwood, 1967), presumably because the former assists implicit grouping. In the present case, suppose that adjacently presented items will be grouped, but only if they contain a common node (e.g., SUR → PIL with FAM → PIL in the levels-up and levels-down conditions, and SUR → PIL with PIL → CUF in the paths-up condition). In this example, it would be more beneficial to group SUR with FAM than with CUF because SUR and FAM have something else in common, i.e., neither is a correct response to any stimulus node. This hypothesis can also be extended to implicit grouping on the basis of reinforcement frequency (Runquist, 1965). A mnemonic group consisting of the upper subtree (levels-down condition) would be more useful than one consisting of a lower subtree (subtree, levels down) because only the former consists precisely of those nodes that are correct responses to one or another stimulus.

The present results, taken together with findings from other studies, suggest several directions for future research. One might query *S*s at the end of the experiment as to the tree structure. Or it might be determined which conditions—nonsense names, random assignment of names to nodes, or testing by recognition methods—accounted for the failure to find dependencies among component learning processes. With natural language hierarchies in which the node names are words, the tendency to learn the overall list structure apparently increases if the entire tree is displayed at once instead of presenting its parts serially. This effect is larger if testing is by recall rather than recognition (Bower et al., 1969). However, in preliminary research using the present paradigm but testing by recall instead of recognition, the proportions of correct responses have been found to be very low. Finally, on the basis of Mandler and Cowan's (1962) finding, hierarchies containing equal numbers of associations to be learned but differing in graph structure will probably be learned at different rates.

REFERENCES

Battig, W. F. Evidence for coding processes in "rote" paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 177–181.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 323–343.

Deese, J. *The structure of associations in language and thought.* Baltimore: Johns Hopkins University Press, 1965.

Mandler, G. Association and organization: Facts, fancies, and theories. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Mandler, G., & Cowan, P. A. Learning of simple structures. *Journal of Experimental Psychology*, 1962, **64**, 177–183.

Miller, G. A. Psycholinguistic approaches to the study of communication. In D. L. Arm (Ed.), *Journeys in science: Small steps, great strides.* Albuquerque: University of New Mexico Press, 1967.

Norman, D. A., & Waugh, N. C. Stimulus and response interference in recognition-memory experiments. *Journal of Experimental Psychology*, 1968, **78**, 551–560.

Runquist, W. N. Order of presentation and number of items as factors in paired-associate verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 535–540.

Wood, G., & Underwood, B. J. Implicit responses and conceptual similarity. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 1–10.

(Received February 28, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 290–296

RESPONSE FEEDBACK AND VERBAL RETENTION[1]

JACK A. ADAMS,[2] JOHN S. McINTYRE,[3] AND HOWARD I. THORSHEIM[4]

University of Illinois

A closed-loop theory of verbal learning was tested, where it is assumed that the sensory feedback from a response lays down the reference against which the correctness of future occurrences of the response is tested. A short-term memory (STM) paradigm was used, with consonant tetragrams as items. Experimental groups had interference with audio feedback, tactual-proprioceptive feedback, or both in acquisition and were compared with a control group having standard, unimpaired learning. Recall was without interference for all groups. It was hypothesized that feedback interference should affect retention, with experimental groups having poorer recall than the control group. The data supported the hypothesis.

Adams (1967, Ch. 10; 1968) has suggested a closed-loop hypothesis for learning that uses two kinds of habit: the memory trace, which is the habit bond between stimulus and response, and the perceptual trace, which is a sensory habit and a consequence of stimulation from the environment or response feedback. To illustrate for a verbal response, rehearsal defines the correct response and establishes a memory trace, which is the associative agent that fires the response when the eliciting stimulus is presented, and the proprioceptive, tactual, and auditory feedback stimuli from the response are the sensory agents that lay down perceptual traces. It is these perceptual traces that are the internal references for response recognition against which the sensory feedback from subsequent responses are compared, error assessed, and the judgment about response correctness is made. In the same way, environmental stimuli lay down a perceptual trace as the reference for subsequent recognition of them.

An implication of the hypothesis is that interference with one or more of the feedback loops during learning should attenuate the reference mechanism based on it, with the result that *S* should be less able to judge the correctness of his response. Recall accuracy should be lower because *S* with a sufficient memory trace for the correct response will have an attenuated reference mechanism and a lowered confidence in the response he intends to make. Being unable to fully verify the correct response, he will tend to substitute another and produce more errors.

METHOD

Experimental design.—The essentials of the Peterson and Peterson (1959) procedures for short-term verbal retention were used, with consonant tetragrams chosen for their low acoustic similarity (Adams, Thorsheim, & McIntyre, 1969) as the materials. The four letters of an item were presented visually and simultaneously, and *S* read the letters aloud while they were showing. Each *S* learned and recalled 20 items, with 4 items randomly assigned to each of five retention intervals: 0, 5, 10, 15, and 30 sec.

Four groups were used, one control group and three experimental groups. The treatment for each experimental group was interference with one or two sensory feedback loops when the response was spoken in acquisition. The control group had the standard procedure for acquisition in short-term memory studies, which, of course, is no interference. All groups had their retention interval activity and recall without interference. The expectation was that impaired reference mechanisms in experimental *S*s would lower recall relative to controls.

The three experimental groups were IAF, IPF, and IAPF. The *S*s of Group IAF (impaired auditory feedback) wore earphones (as a control, all *S*s of all groups wore the headset throughout) that delivered 110-db. white noise to eliminate the sound of *S*'s own voice as he spoke the tetragram aloud when it was presented. Pilot testing had established that 110 db. effectively eliminated self-hearing of the normal speaking voice. The noise came on 1 sec. before the item appeared and continued 1 sec. after it was removed. As a means of altering the tactual-proprioceptive cues from learning to recall, an *S* of Group IPF (impaired proprioceptive feedback) clamped his tongue between his teeth and held a deep breath when he spoke the letters of an item during acquisition (adapted from a procedure used by Colvin & Meyer, 1909). Group IAPF (impaired auditory and proprioceptive feedback) had a compound of the treatments used on Groups IAF and IPF, with both white noise and the tongue–breath treatments used together. The control group was labeled NIF (no impaired feedback). If the hypothesis is empirically true, Groups IAF and IPF should have poorer recall than Group NIF, and Group IAPF should be poorer than both IAF and IPF.

---

[1] This research was supported by Grant USPH MH-12022 from the United States Public Health Service. Thomas O. Nelson assisted in the experiment.

[2] Requests for reprints should be sent to Jack A. Adams, Department of Psychology, University of Illinois, Urbana, Illinois 61801.

[3] Now at the University of Manitoba.

[4] Now at St. Olaf College.

As an added control, two of the four items at each retention interval for the three experimental groups were randomly assigned as standard, noninterference control items, the same as Group NIF. A possibility was that an interference treatment might be a nonspecific factor like annoyance that could pervade all items, not just treatment items. Conceivably, the recall decrements expected because of degraded reference mechanisms could occur because of some other consequence of the treatments. Therefore, within-group control items, temporally near treatment items, might be expected to reflect a general state change in *S*.

Two basic random orders of the 20 items were used, and there were two versions of each order to achieve counterbalancing of control and treatment items. The two items of an interval that were treatment items in one version were control items in the other. The two versions of the two orders resulted in four item sequences for the experimental groups, and an equal number of *S*s of each group were randomly assigned to each sequence. Group NIF had no special treatment items, so only the two basic random orders were used, with an equal number of *S*s in each order.

Apparatus.—The presentation device was a Lafayette memory drum (Model 2303A) modified for timing by a punched tape reader placed outside the experimental room.

Experimental procedures.—The sequence of events on a trial was as follows.

1. Ready signal for 5 sec. For Group NIF this was always the word "Ready." As their ready signals, the experimental groups had special instructions for the treatment requirements of the item to follow. The ready signal for control items was the word "Standard." For treatment items, Group IAF had the word "Noise" as the ready signal, Group IPF had "Tongue, deep breath," and Group IAPF had "Noise, tongue, deep breath." The *S*s of Groups IPF and IAPF used this 5-sec. period to place the tongue between the teeth and take a deep breath.

2. Item presentation for 3 sec. and letters read aloud. As might be expected, *S*s of Groups IPF and IAPF articulated their treatment items poorly, but then the purpose of the tongue–deep-breath procedure was to alter the proprioceptive-tactual feedback from learning to recall.

3. Retention interval of 0–30 sec. To minimize rehearsal in the retention interval, *S*s read random digits as fast as they could. The digits were presented 15 to the line in the window of the drum, and a new line appeared every 3 sec. Since *S*s read digits no faster than 3/sec, they never finished a line and had no free time for item rehearsal before the next line appeared.

4. The *S* had 10 sec. for recall, and the recall cue was the words "Recall-Rating." Recall was aloud, and *E* recorded the responses on his data sheet. The *S* always had to report four letters, even if he had to guess, as a means of preventing him from inhibiting letters about which he was unsure. After his recall, *S* reported to *E* the confidence he had in his response on a 1–5 scale, ranging from "I'm absolutely sure that my response is wrong" (rating of 1) to "I'm absolutely sure that my response is correct" (rating of 5). For reference, the five scale values and their definitions were on a card on the table in front of *S*.

5. Ten seconds were allowed for *S*'s oral report of how he learned the item. Previous studies have documented that *S*s often use idiosyncratic associative devices, called natural language mediators (NLM), in both long-term (Adams & McIntyre, 1967; Montague, Adams, & Kiess, 1966) and short-term retention (Adams et al., 1969; Groninger, 1966; Kiess, 1968). All these studies have shown that NLMs induce or are a correlate of higher recall than rote learning, and reports of them were taken in this study so that the feedback theory could be assessed with respect to them. The *S* reported his NLM in full, but *E* recorded only whether an association was present or absent. Absence of an NLM was tallied as rote learning.

The instructions included two demonstration items that familiarized *S*s with the general procedures and, in the case of *S*s in experimental groups, acquainted them with the interference treatments. The *E* counseled against the common tendency to speak louder in the presence of white noise because a compensatory device of this sort would have allowed *S* to hear his own voice and would have undermined the effectiveness of the noise treatment. The instructions gave no emphasis to either mediated or rote learning, but described them only as ways of learning; *S*s were told to conscientiously report their method each time.

Subjects—The *S*s were men and women university students from an introductory psychology course. Two hundred and forty *S*s were used, with 60 randomly assigned to each group.

RESULTS

Percentage of items correct (mediated plus rote).—All letters had to be in their correct serial positions for an item to be scored as correct, and Fig. 1 shows the per-

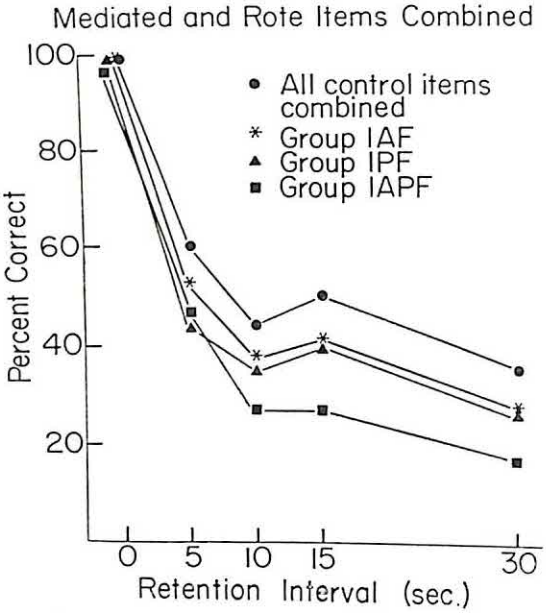


FIG. 1. Percentage of items correct as a function of retention interval.

centage of items correct for experimental groups and for all control items in the study pooled. The within-group control items of the experimental groups had the same performance level as Group NIF, F (3, 236) = .37, $p > .05$, which justifies the pooling.

Of more importance than the now commonplace rapid forgetting over 30 sec. is that theoretical expectations are confirmed in Fig. 1. Using number of correct items as the score for S in the statistical testing, and pooling over intervals, treatment items for each experimental group had significantly poorer performance than their control items: Group IAF, F (1, 59) = 10.7, $p < .01$; Group IPF, F (1, 59) = 12.01, $p < .01$; Group IAPF, F (1, 59) = 23.3, $p < .01$. The order of the curves for treatment conditions conforms to the number of feedback loops receiving interference, with Group IAPF (two loops) having lower recall than Groups IAF and IPF (one loop), which have equivalent recall. Essentially, the decrements for Groups IAF and IPF additively produce IAPF's decrement.

Mediation.—It is established that mediated items have higher recall than rote ones, and this study was no exception. For all items of all groups combined, 73% of mediated items were correct compared to 43% of rote items.

An effect of feedback interference could be less NLM production, thereby lowering recall and causing the ordering of treatment curves in Fig. 1 because correct recall is related to NLM occurrence. Percentage of NLMs for both correct and incorrect items taken over all retention intervals for Groups NIF, IAF, IPF and IAPF were, respectively, 38, 36, 36, and 40 for control items and 41, 30, 32, and 31 for treatment items. (All items were control items for Group NIF, but the items can be split into the same two sets as as other groups for purposes of comparison.) There is a tendency for treatments to reduce the level of mediation for the experimental groups, and this trend received some statistical support. Using number of NLMs as the score for S, the differences between control and treatment items for Groups IAF, F (1, 59) = 10.7, and IAPF, F (1, 59) = 11.1, were both significant ($p < .01$). The same comparison for Groups IPF, F (1, 59) = 3.37, and NIF, F (1, 59) = 1.65, failed significance at the .05 level. But there is more to treatment effects than NLM frequency. When NLMs were eliminated as a factor and rote items alone were considered, the treatment items for experimental groups had significantly less percentage correct than control items, F (2, 111) = 4.8, $p < .05$. This analysis was over all retention intervals and used Ss who had two or more rote items. The n's for Groups IAF, IPF, and IAPF were 37, 41, and 36, respectively, for this particular test.

Confidence ratings.—An analysis was made of the confidence ratings for correct responses (mediated plus rote) taken over all intervals, with mean confidence rating for S's items as his score. Treatment items for experimental groups had significantly lower ratings than control items, F (1, 177) = 4.0, $p < .05$, although the three groups did not differ appreciably among themselves, F (2, 177) = .15, $p > .05$. The same trend was found for confidence ratings of errors (mediated plus rote) made by experimental groups. Treatment items were rated significantly lower than control items, F (1, 177) = 23.4, $p < .01$, but the groups did not

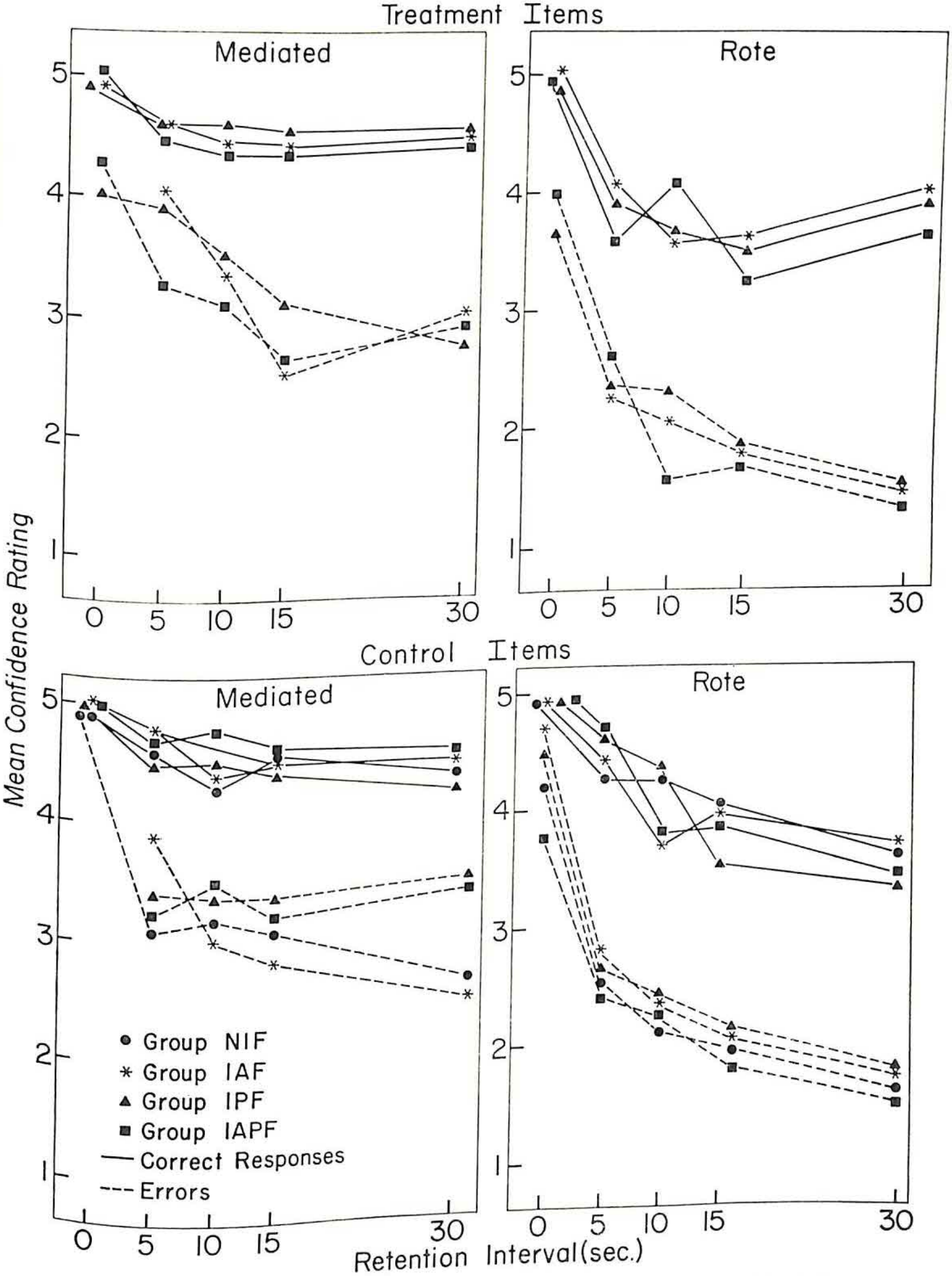


FIG. 2. Mean confidence rating for correct and incorrectly recalled items. (A missing point for a curve means insufficient items ($N < 2$) on which to compute a mean.)

differ among themselves, $F\ (2,\ 177) = 2.1$, $p > .05$.

Figure 2 presents mean confidence ratings for items correctly and incorrectly recalled, mediated and rote items, and treatment and control items as a function of retention interval. For both correct and incorrect items, the mediated items have higher confidence ratings than rote items. Correct items, by definition, have full retention and yet their confidence ratings decrease as the retention interval increases, with the drop being greater for rote items. The confidence ratings for errors are much lower than for correct responses, and they decrease more precipitously as a function of retention interval. Again the rate of loss is greater for rote than mediated items.

Confidence rating and degree of correctness (zero, one, two, or three letters correct) were positively correlated, and one of the reasons for the sharp drop in errors over the retention interval was an increase in errors with a low degree of correctness as the interval increased. The lowest and highest degrees of correctness were compared at the 5- and 30-sec. retention intervals (the 0-sec. interval had too few errors to use meaningfully); errors with the highest degree of correctness decreased as the interval increased, while errors of the lowest degree increased. But this was not the only reason that confidence in errors declined. With degree of error correctness held constant, confidence was less at the 30- than the 5-sec. interval.

Rote error items were rather evenly divided among the four degrees of correctness, but mediated errors tended to be more correct, with two and three letters correct occurring more frequently. This is consistent with the established finding that NLMs are a correlate of higher retention than rote learning. That mediated errors are more correct than rote errors is undoubtedly a factor in their higher confidence rating, but the explanation is not so simple. With degree of correctness held constant, mediated responses had higher confidence ratings than rote responses. An *S* is more sure of his mediated responses.

DISCUSSION

Comfort or contradiction was sought in the literature on the effects of noise and of articulatory interference on verbal learning and retention, but neither was found. Studies of STM using white noise were of potential relevance and had mixed purposes and findings. Two recent studies with an STM paradigm (Dallett, 1964; Murdock, 1967) found an effect of noise, but audio presentation of items was used so the influence was on the quality of the input, in contrast to the present study, which guaranteed the input by using visual presentation and sought only to degrade response feedback. The most pertinent experiment was by Murray (1965). His STM design had visual presentation and immediate recall of eight consonants, and one of his conditions had acquisition with noise and recall without noise. The trend was toward a recall decrement, but the loss failed statistical significance. Studies that attempted to manipulate articulation (Barlow, 1928; Colvin & Meyer, 1909; Henmon, 1912; Mould, Treadwell, & Washburn, 1915; Smith, 1896; Underwood, 1964) did not use STM paradigms, and their array of positive and negative findings did not focus on a useful theme for the present experiment.

Trying to be our own best critics, we saw three main criticisms using conventional variables that might be leveled against our feedback interpretation:

1. *Distraction.*—Under conditions of feedback interference, *S*s paid less attention to the items, with the result that the input for experimental groups was not of the strength or quality of control items. As a rebuttal to this contention, all *S*s pronounced the letters aloud, thereby guaranteeing a measure of attention. That effective input occurred is shown by the fact that all groups were about the same and close to 100% correct on the immediate retention test.

2. *Associative strength.*—Related to the distraction argument is that interference did not affect input, but did influence the amount of covert rehearsal that followed reading the letters aloud, with the outcome of less associative strength and more forgetting. All groups being essentially 100% correct at the immediate retention test could be an argument for equal associative strength. The argument is not a strong one, however, because 100% is a ceiling that obscures group differences, and the authors admit the possibility of unequal as-

sociative strength as a criticism that is as hard to refute as it is to prove. "Associative strength" as it is customarily used is the same as strength of memory trace, and the validity of the present conclusions turns on the effectiveness of the procedures for equating memory traces while at the same time impairing the integrity of sensory feedback and the perceptual traces based on it. Equating items for associative strength appears to be more complex than is conventionally assumed (Adams, 1967, pp. 140–143), so no easy resolution of this problem seems apparent.

3. *Annoyance.*—A subjective term, annoyance is hard to fit into theoretical criticism. Attempting a working definition nevertheless, annoyance might be seen as avoidance motivation that could reduce the frequency of covert rehearsal and lower associative strength (Wickens & Simpson, 1968) or, alternatively, lower all reactivity, both learning and recall. Being vague, annoyance cannot be solidly refuted with evidence or counterarguments, but the authors nevertheless tried to anticipate this criticism with the within-group control items for the experimental groups. Subjectively, annoyance has time persistence, and it was reasoned that it should be revealed by control items intermingled with treatment items. That performance of experimental *S*s on control items did not differ from that of Group NIF is evidence against annoyance as a variable.

Remarks on the positive side for the theory are in order now that criticisms have been acknowledged, and the most encouraging is the percentage of correct functions in Fig. 1, with the curves for treatment items below control items and the order of treatment items for experimental groups in line with theoretical expectations. Furthermore, confidence rating is an empirical indicant of uncertainty about the current response, and lower confidence for treatment than control items suggests less adequate perceptual traces for verifying the response. The *S*s were less sure of correct treatment items when they gave them, as well as errors. The lower confidence in errors could mean that *S*s sometimes covertly tested the correct response based on a strong memory trace and, being unable to verify it, changed it to another and wrong response at recall. This is, in effect, a guess at one or more letters and is a circumstance for lower confidence.

Fully correct responses, by definition, are 100% correct at each retention interval and have a fully functioning memory trace, yet their confidence ratings declined as the retention interval increased (see Fig. 2). If this trend continued over longer time intervals than used here, and there is no reason to believe that it would not, confidence might be expected to get low enough to cause *S* to reject his response for another (a guess) or inhibit it and cause an omission to occur (although in this experiment, omissions were not allowed). For a correct response this would produce an error, and measured retention would decrease. Confidence ratings for errors of a given degree of correctness similarly declined, suggesting that the degree of error could be expected to increase also. This decline in confidence over the retention interval can be taken as evidence that the perceptual traces are being weakened by forgetting processes (Adams, 1967, pp. 302–304). The experimental psychology of memory has dwelt on the memory trace that is strengthened in learning and weakened over time, but feedback theory says that change in the perceptual trace is equally deserving of concern. Put another way, experimental psychology has oversimplified what is learned by dwelling solely on the memory trace and thus has oversimplified forgetting. An implication of this point of view is that a profitable line of research for memory is the complexities of the learning process, for we cannot understand what is forgotten unless we know what has been learned in the first place.

REFERENCES

ADAMS, J. A. *Human memory.* New York: McGraw-Hill, 1967.

ADAMS, J. A. Response feedback and learning. *Psychological Bulletin,* 1968, **70**, 486–504.

ADAMS, J. A., & MCINTYRE, J. S. Natural language mediation and all-or-none learning. *Canadian Journal of Psychology,* 1967, **21**, 436–449.

ADAMS, J. A., THORSHEIM, H. I., & MCINTYRE, J. S. Item length, acoustic similarity, and natural language mediation. *Journal of Experimental Psychology,* 1969, **80**, 39–46.

BARLOW, M. C. The role of articulation in memorizing. *Journal of Experimental Psychology,* 1928, **11**, 306–312.

COLVIN, S. S., & MEYER, E. J. The development of imagination in school children and the relation between educational types and retentivity of material appealing to various sense departments. *Psychological Review,* 1909, **11**(1, Pt. 2).

DALLETT, K. M. Intelligibility and short-term memory in the repetition of digit strings. *Journal of Speech and Hearing Disorders,* 1964, 7, 362–368.

GRONINGER, L. D. Natural language mediation and covert rehearsal in short-term memory. *Psychonomic Science,* 1966, **5**, 135–136.

HENMON, V. A. C. Relation between mode of presentation and retention. *Psychological Review,* 1912, **19**, 79–96.

KIESS, H. O. Effects of natural language mediators on short-term memory. *Journal of Experimental Psychology,* 1968, **77**, 7–13.

MONTAGUE, W. E., ADAMS, J. A., & KIESS, H. O. Forgetting and natural language mediation. *Journal of Experimental Psychology,* 1966, **72**, 829–833.

MOULD, M., TREADWELL, L., & WASHBURN, M. F. Minor studies from the psychological laboratory of Vassar College: XXVII. The influence of suppressing articulation on the favorable effect of distributing repetitions. *American Journal of Psychology,* 1915, **26**, 286–288.

MURDOCK, B. B., JR. The effects of noise and delayed auditory feedback on short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 737–743.

MURRAY, D. J. The effect of white noise upon the recall of vocalized lists. *Canadian Journal of Psychology,* 1965, **19**, 333–345.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology,* 1959, **58**, 193–198.

SMITH, T. L. On muscular memory. *American Journal of Psychology,* 1896, **7**, 453–490.

UNDERWOOD, B. J. Articulation in verbal learning. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3**, 146–149.

WICKENS, D. D., & SIMPSON, C. K. Trace cue position, motivation, and short-term memory. *Journal of Experimental Psychology,* 1968, **76**, 282–285.

(Received March 3, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 297–300

AMOUNT AND LOCUS OF STIMULUS-RESPONSE OVERLAP IN PAIRED-ASSOCIATE ACQUISITION[1]

DOUGLAS L. NELSON[2] AND RICHARD M. GARLAND

University of South Florida

Each *S* acquired a single list of three-letter-word paired associates. Stimuli shared zero, first, middle, last, first-and-middle, first-and-last, or middle-and-last letters with their responses. On the basis of data suggesting that some ordinal positions within the word sequence are more critical determinants of word processing than others, differential acquisition rates for these conditions were predicted. Two alternative explanations of the findings were discussed, one assuming that the informational structure of the language has sensitized speakers of the language to attend to high-information parts of words, and one assuming that words are processed as serial sequences.

Contemporary research has shown that initial and final letters are more influential determinants of word processing than are middle letters and that initial are more critical than final letters. Bruner and O'Dowd (1958) presented nouns containing typographical reversals of letters in either the beginning, middle, or end. Disruption of tachistoscopic recognition performance was greatest with a reversal in the initial part of the word, next greatest at the end of the word, and least disruption was found when the error was localized in the middle of the word. Following a single presentation of six- or nine-letter words, Horowitz, White, and Atwood (1968) showed fragments of these words in a recall test. Regeneration of the words of the list was highest when beginning fragments were shown, next highest when end fragments were shown, and lowest when middle portions were shown. Similarly, given a word containing a deleted letter, correct identification has been shown to increase as the missing letter is moved from initial to final and medial positions (Carson, 1961). The initial–final difference has also been found using native speakers of the language and Dutch and French words (Broerse & Zwaan, 1966; Oleron, 1966).

Differential sensitivity to word parts may account for findings obtained in experiments in which there were variations in the ordinal position of letters shared by the stimuli (i.e., identical letters were located in the first, second, etc., positions reading the trigram from left to right). When discriminations between stimuli were required as in paired-associate and serial recall learning, difficulty of acquisition increased in the order middle, last, and first for single-locus conditions and in the order middle-and-last, first-and-middle, and first-and-last for dual-locus conditions (Nelson, 1969; Nelson & Rowe, 1969). A similar ordering of difficulty for dual-locus conditions also has been found when the paired-associate stimuli were nonsense syllables (Richardson & Chisholm, 1969; Runquist, 1968), suggesting that word-encoding habits may transfer to these stimuli. When shared-letter stimuli could be grouped together as in free recall learning, the orders of difficulty by identity condition were reversed (Nelson, 1969).

The purpose of the present experiment was to manipulate the number and ordinal positions of letters shared by paired-associate stimuli and their responses. The principal hypothesis was that the greater the compatibility between word-processing habits and the location of the stimulus–response overlap, the faster a given list should be acquired. For example, if initial relative to final letters are more critical determinants of word encoding, then stimulus–response pairs containing first-letter overlap should be learned faster than pairs with last-letter overlap. Thus, in terms of increasing ease of acquisition, the order of overlap conditions should be zero, middle, last, first, middle-and-last, first-and-middle, and first-and-last. The only other available evidence on the variable of location of overlap has been reported by Thompson and Fritzler (1968). Ease of acquisition for nonsense syllable pairs increased in the order zero, last, and first, although only the difference between zero and first was statistically reliable.

---

[1] This research was supported by the University of South Florida Research Council and by Grant MH 16360-01 to the first author from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Douglas Nelson, Department of Psychology, University of South Florida, Tampa, Florida 33620.

METHOD

Verbal material.—Each *S* acquired a single list of five paired-associate trigrams in an individual session. Two sets of lists were constructed. Within each set, there was a single response list common to seven stimuli lists. The stimuli within a given list set had either zero, first, middle, last, first-and-middle, first-and-last, or middle-and-last letters in common with their responses. For example, under the first-letter locus condition, the pairs of one list set were PIG-*pan*, ROW-*red*, SAY-*sit*, CUP-*cob*, and HEX-*hum*; for first-and-middle, PAY-*pan*, REX-*red*, SIP-*sit*, COW-*cob*, and HUG-*hum*. Within the response lists and in each of the stimulus lists, there was no letter identity within a given ordinal locus. When stimulus identity was manipulated with respect to a given location(s), remaining letter(s) were different. When possible, these letters were chosen from letters not used in constructing the remaining stimuli and responses.

Since the items were relatively high-frequency words, differences in rates of acquisition between conditions might have reflected degree of prior association rather than amount and location of stimulus–response overlap. To circumvent this potential interpretative difficulty, associational norms were obtained. In two 30-min. sessions a total of 350 three-letter words were read aloud, spelled, and pronounced aloud again to four different groups of 25 students each. Word order was unsystematically randomized for each group. In response to each word read, each student wrote down the first three-letter word that came to him on a numbered sheet provided by *E*. The frequency of each of the responses given to each of the stimulus words was tabulated and this information was used in constructing the pairs used in the present experiment. If more than 5% of the individuals responding gave a particular response to a given stimulus, then that stimulus–response pair was eliminated as a possible candidate for the list. Thus, not more than 5% of the sample tested gave *pan* as a response to PIG or *pig* as a response to PAN. In this manner degree of prior association was equated across identity conditions.

Procedure.—Each paired-associate list was practiced by the anticipation method for a total of 15 trials following the initial presentation. On test trials, *S* was required to pronounce each stimulus aloud before attempting anticipation of the response. Slides of each stimulus and each stimulus–response were automatically projected via a Kodak Carousel slide projector driven by an auxiliary timer at a 1.5:1.5-sec. rate. There was no intertrial interval. Sequences of pairs were presented in eight unsystematically varied orders, with the restriction that no pair was ever presented twice in succession. The sequence of each of these eight orders was changed for each *S* in each condition. To insure understanding of the instructions, all *S*s learned a practice list of two pairs (names-verbs) before undertaking the experimental task.

Subjects.—With half of the *S*s assigned to each list set, there were seven groups of 20 *S*s, all of whom were naive to verbal learning experiments. All *S*s were drawn from an introductory psychology course and received course credit for participation. The *S*s were assigned to conditions in order of appearance in blocks of 14, with 1 *S* from each condition per block. Order of conditions within blocks was determined by a table of random numbers.

RESULTS

A mixed-model analysis of variance (Myers, 1966) of errors (omission plus commission) showed that identity conditions, trials, and the interaction between these sources were significant, the latter two sources remaining significant following the Greenhouse and Geisser (1959) degrees-of-freedom adjustment for repeated measures variables. List set and all interactions with this variable were not significant. Table 1 presents the means of errors to the acquisition criterion for each of the overlap conditions. Fisher's least significant difference was 1.01, with the 5% level as the criterion of significance. Each mean was significantly different from each other mean except means for first-and-middle and middle-and-last. These data show that ease of acquisition increased as amount of stimulus–response overlap increased from zero to one and two letters. The significant Identity Conditions × Trials interaction reflected the disproportionately faster acquisition rates for the dual treatments relative to single-locus and zero conditions. Within single-letter-overlap conditions, acquisition ease increased in the order middle, last, and first, and within dual-overlap conditions, the order was first-and-middle, middle-and-last, and first-and-last. With the exception of the middle-and-last condition, this pattern of mean differences was consistent with the hypothesis of dif-

TABLE 1

Mean Errors, Mean Percentage of Misplaced Responses and Mean Percentage of Extralist Intrusions for Each Condition

Item	Overlap conditions						
	Zero	First	Middle	Last	First-and-middle	First-and-last	Middle-and-last
Mean errors	23.40	11.75	19.80	17.10	7.35	5.45	6.90
Mean % misplaced responses	14.55	.83	12.85	9.45	.00	1.05	1.87
Mean % extralist intrusions	2.40	7.61	3.09	3.39	9.55	11.17	11.37

ferential sensitivity to beginning, ends, and middles of words.

Mean misplaced responses and extralist intrusions as percentages of total errors are presented in Table 1. A single-factor analysis of variance of percentage of misplaced responses yielded a significant identity conditions effect. A subsequent orthogonal comparison indicated that the proportion of misplaced responses significantly declined in the order zero, single, and dual locus. Although a similar analysis of variance of percentage of extralist intrusions revealed an insignificant *F* ratio for conditions, a subsequent orthogonal comparison indicated that proportion of extralist intrusions significantly increased in the order zero, single, and dual locus. These patterns of commission errors suggest that amount of overlap facilitated response retrieval by constraining the size of the relevant response set. When an extralist intrusion occurred, it generally shared a letter, if not two, with the stimulus presented.

Discussion

The differential sensitivity to ordinal positions of letters within words observed in the present study may in part be reflecting an adaptation by speakers of the language to the informational structure of words. Beginning and end letters carry more information than middle letters, and beginning letters carry more than end letters (Carson, 1961; Nelson & Rowe, 1969). In other terms, the frequency distribution of words beginning with each of the 26 letters is more uniformly rectangular than the distributions for middles and ends, etc. Thus, presentation of stimuli which overlap with their responses in the initial letter position provide more information about their responses than stimuli which overlap in middle or last positions. Similarly, stimulus–response overlap in first-and-last letter positions produces more constraint on the responses given the stimuli than overlap in the first-and-middle or middle-and-last positions. If the response retrieval process can be considered analogous to a search through memory for the correct alternative, the more information the stimulus provides about the response, the more likely the search is to terminate with the correct alternative (e.g., Ceraso, 1967; Shiffrin & Atkinson, 1969). In the present experiment, amount and location of overlap serves to vary the size of the relevant response set which must be searched before the correct response is located. The smaller the set of alternatives through which the search must be conducted, the greater the likelihood that the correct instead of an incorrect response will be retrieved. The results of the present experiment have been generally consistent with this view. The lack of a difference between first-and-middle and middle-and-last conditions may be a result of the very small differences between these conditions in information values, which are, respectively, 7.06 and 6.89 bits for three-letter words (Nelson & Rowe, 1969). The most serious discrepancy, however, arises in comparison of information values for single-letter positions and performance in single-locus identity conditions. First and third letters are more similar in information value (4.46 vs. 4.21 bits) than are second and third letters (3.35 vs. 4.21 bits). However, performance within the last-letter identity condition is more similar to middle than to the first-letter identity condition.

A viable alternative interpretation of data indicating differential sensitivity to word parts assumes that the processes involved in word encoding are similar to those involved in serial anticipation or serial recall learning (Harcum, 1967; Horowitz, 1969). In this view, words are stored and/or retrieved from memory as serial sequences having definite starting points and ends. Accordingly, explanations of the serial position effect derived from studies of serial anticipation learning also would account for the recall of word sequences. Thus, differential attention to or recall of various ordinal positions within the word sequence may be a result of differential interletter interference in various parts of the sequence or to the perceptual saliency or distinctiveness of the beginning and end of the sequence (e.g., Bewley, Nelson, & Brogden, 1969; Lippman & Denny, 1964; Murdock, 1960). In the present experiment the acquisition of each pair would be analogous to the simultaneous practice of two serial lists which overlap to varying degrees, one "list" represented by the stimulus word and one by the response word. Providing initial portions apparently provides greater accessibility to the sequence than medial or terminal portions, and terminal provide slightly better access than medial portions. The nature of the mechanism producing this differential accessibility is not clear, but one possibility is that beginnings and, to a lesser extent, ends provide perceptual anchoring points for encoding the word. The particular salience of the beginnings of words may be a result of left-right reading habits or of the way in which the words are organized in storage.

At the present level of descriptive analysis, either the hypothesis of sensitivity to the informational structure of language or the hypothesis that words are encoded as sequential patterns provides a reasonable explanation of the findings of the present investigation. A decision between these alternatives at this level of description may prove difficult. The differential informational loading for locus of letters within words and for the locus of words within sentences (e.g., Aborn, Rubenstein, & Sterling, 1959) may lead to the use of general sequential search strategies which transfer to other serial tasks. Alternatively, through the course of language development, most of the information may be located at the beginning and end of the sequence because of anchoring or serial position effects. In this case information is provided in that part of the sequence where it is optimal for message transmission (Zipf, 1949).

REFERENCES

Aborn, M., Rubenstein, H., & Sterling, T. D. Sources of contextual constraint upon words in sentences. *Journal of Experimental Psychology,* 1959, **57**, 171–180.

Bewley, W., Nelson, D. L., & Brogden, W. J. Effect of cut alternation for ordinal position on acquisition and serial position curve form. *Journal of Experimental Psychology,* 1969, **79**, 445–451.

Broerse, A. C., & Zwaan, E. J. The information value of initial letters in the identification of words. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 441–446.

Bruner, J. S., & O'Dowd, D. A note on the informativeness of parts of words. *Language and Speech,* 1958, **1**, 98–101.

Carson, D. H. Letter constraints within words in printed English. *Kybernetik,* 1961, **1**, 46–54.

Ceraso, J. The interference theory of forgetting. *Scientific American,* 1967, **217**, 117–124.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. *Psychometricka,* 1959, **24**, 95–112.

Harcum, E. R. Parallel functions of serial learning and tachistoscopic pattern perception. *Psychological Review,* 1967, **74**, 51–62.

Horowitz, L. M. Sequential learning, associative symmetry, and thought. In J. F. Voss (Ed.), *Approaches to thought.* Columbus, Ohio: Charles E. Merrill, 1969.

Horowitz, L. M., White, M. A., & Atwood, D. W. Word fragments as aids to recall: The organization of a word. *Journal of Experimental Psychology,* 1968, **76**, 219–226.

Lippman, L. G., & Denny, M. R. Serial position effect as a function of intertrial interval. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3**, 496–501.

Murdock, B. B. The distinctiveness of stimuli. *Psychological Review,* 1960, **67**, 16–31.

Myers, J. L. *Fundamentals of experimental design.* Boston: Allyn & Bacon, 1966.

Nelson, D. L. Information theory and stimulus encoding in free and serial recall: Ordinal position of formal similarity. *Journal of Experimental Psychology,* 1969, **80**, 537–541.

Nelson, D. L., & Rowe, F. A. Information theory and stimulus encoding in paired-associate acquisition: Ordinal position of formal similarity. *Journal of Experimental Psychology,* 1969, **79**, 342–346.

Oleron, P. The role of relative indeterminacy in word initials and finals: A study in word identification. *Language and Speech,* 1966, **9**, 201–206.

Richardson, J., & Chisholm, D. C. Transfer of cue selection based on letter position. *Journal of Experimental Psychology,* 1969, **80**, 299–303.

Runquist, W. N. Formal intralist similarity in paired-associate learning. *Journal of Experimental Psychology,* 1968, **78**, 634–641.

Shiffrin, R. M., & Atkinson, R. C. Storage and retrieval processes in long-term memory. *Psychological Review,* 1969, **76**, 179–193.

Thompson, C. P., & Fritzler, D. E. Paired-associate acquisition: Some effects of inter- and intrapair similarity. *Journal of Experimental Psychology,* 1968, **78**, 107–112.

Zipf, G. K. *Human behavior and the principle of least effort.* Cambridge, Mass.: Addison-Wesley, 1949.

(Received March 3, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 301–309

FUNCTIONING OF IMPLICIT ASSOCIATIVE RESPONSES IN MEDIATED TRANSFER[1]

PHEBE CRAMER[2]

University of California, Berkeley

Using an A′-B, A-C paradigm, a $3 \times 3 \times 2$ factorial experiment was designed that varied the strength of implicit associative responses (IARs), the degree of first-list learning, and the locus of mediation. The results from 18 independent groups of *S*s ($N = 16$/group) indicated that transfer was a function of the interaction of these three variables. With increasing degrees of original learning, transfer became positive for the unidirectional paradigms, but remained negative for bidirectional paradigms. The results are interpreted as supporting three hypotheses: (*a*) Presentation of a stimulus word results in the automatic elicitation of IARS of differing strengths in the associative hierarchy, any one of which may function to mediate transfer; (*b*) in a transfer task, such IAR elicitation occurs both during original list (OL) and test list (TL) learning; (*c*) under certain conditions, if the elicitation of such IARs produces interference, *S* is able to inhibit their occurrence.

The conception of mediational processes based on implicit associative responses (IARs) has been shown to have considerable utility for predicting verbal behavior and learning (cf. Kjeldergaard, 1968). As part of this general area of study, the role of mediation in paired-associate (PA) transfer has been the topic of a series of investigations (cf. Keppel, 1968). Although there have been some dissenting voices (e.g., Mandler & Earhard, 1965), there is a large body of evidence that clearly points to the positive effects of mediational processes in facilitating transfer over a wide number of paradigms (e.g., Cramer, 1967; Cramer & Cofer, 1960; Horton & Kjeldergaard, 1961).

As in these transfer studies, a large number of other investigations have shown the positive effects of IARs in facilitating learning and memory phenomena. Underwood (1965) has noted that we demonstrate the validity of our assumptions regarding IARs by constructing experimental situations in which a given phenomenon will result if these IARs occur in the way in which we have assumed. The present study is designed to investigate some *negative* transfer effects of IARs that would be predicted from a particular model of mediational processes.

Underwood (1965) has distinguished between two different types of such IARs. The first type (to be designated here as IAR_1) is a word that is an associative response to the stimulus word, as known from natural language word association norms. The second type (IAR_2), which occurs in an experimental situation such as mediated transfer, is the word (or other item) that *S* has learned as a response to a given stimulus in a preceding list. In a previous study (Cramer, 1967), it has been shown that IAR_1 and IAR_2 may differentially affect transfer.

While it is probably most common to think of IAR_1 as the primary associative response to the stimulus, the theory underlying this conception—that IAR_1 is produced by the stimulus properties of the representational response (RR)—also predicts that other responses in the associative hierarchy of the stimulus word will be implicitly elicited. Bousfield, Whitmarsh, and Danick (1958), in their original formulation of the RR and IAR conception, do specifically make this prediction. Indeed, such an assumption is necessary to explain the data of Kimble

---

[1] This report was written at the Institute of Human Learning, which is supported by grants from the National Science Foundation and the National Institutes of Health. The author is indebted to Geoffrey Keppel and Leo Postman for many helpful comments about the paper.

[2] Requests for reprints should be sent to Phebe Cramer, Institute of Human Learning, University of California, Berkeley, California 94720.

(1968) and Underwood (1965), who found that IARs of differing response strengths and hierarchical positions produce recognition errors.

As part of the model of mediational processes underlying the present experiment, it is assumed that any member of an associative response hierarchy can function as an IAR_1 mediator and that the magnitude of the mediated effect will be related to the S-R associative strength of the mediator. One question investigated in this study is the role of IAR associative strength in mediating transfer.

It also is assumed that IARs representing the entire associative hierarchy occur automatically following the presentation of a stimulus word, but that under certain conditions *S*s can and do prevent their occurrence. Intuitively, it is clear that the verbal performance of an individual would be hopelessly handicapped if the occurrence of each verbal utterance were automatically and indiscriminately followed by the elicitation of a hierarchy of IARs. Yet, sometimes in errors of speech and writing, as well as in the reports from experimental *S*s regarding their learning strategy, it seems equally clear that this is what has happened. In situations in which the occurrence of IARs provides a source of interference or confusion, it is assumed that the capacity to inhibit their occurrence is related to the degree of differentiation between the PA stimulus and the interfering IAR. In the present study, the question of whether interfering IARs can be "turned off" as a function of increased differentiation was investigated by varying the degrees of original list (OL) learning.

This conception of the functioning of IARs, as applied to an A′-B, A-C transfer situation, is presented in Fig. 1. During OL learning, *S* learns to associate A′ and B. Leaving aside for the moment the occurrence of IARs during OL, let us consider the processes assumed to occur during test list (TL) learning. If A′ is an associate to A, then when A is presented as a PA stimulus, A′ will occur as an IAR_1 and will serve as a mediating link to Response B, which in turn will interfere with the acquisition of Response C. However, any factor that helps differentiate A′ from A, such as increased OL learning, will facilitate the "turn off" of A′ during TL learning and will thus reduce response interference.

Finally, this study is concerned with the problem of where or when in the course of learning the mediating response occurs. Keppel (1968) has cited, in his review, evidence pointing to test stage mediation, while other investigators (cf. Mink, 1963; Underwood, 1965) have provided convincing evidence that IARs must be elicited during training. This third question is investigated by constructing two different types of A′-B, A-C lists, in which the directionality of the associative relationship between OL and TL varies. A consideration of Fig. 1 in connection with the following will help to explicate the experimental conditions. For both unidirectional and bidirectional paradigms, it is assumed, given $A \rightarrow A'$, that A′ will occur as an IAR_1 on the TL, thereby mediating the occurrence of B as a competing response with C. However, in the bidirectional paradigm, the additional $A' \rightarrow A$ associative relationship is assumed to result in A occurring as an IAR_1 during OL. In this case, A will be associated with B. Subsequently, when A is the TL stimulus, this association mediates the occurrence of B as an IAR_2, which interferes with the acquisition of Response C. The bidirectional paradigm thus has a second source of mediated response interference.

If this model of mediational processes is correct, there should be greater interference when bidirectional lists are used than when unidirectional lists are used. It should be emphasized that this model assumes that IAR_1 (A) is occurring automatically during the OL stage and that B is being established as an IAR_2 to A.

METHOD

Design.—The experiment was designed as a $3 \times 3 \times 2$ factorial, with 18 independent groups. Each *S* learned two lists in an A′-B, A-C paradigm. There were three levels of OL learning (low, moderate, and high), three strengths of associative relationship between the stimulus words of OL and TL (OL was a primary associate, a weak-strength associate, or a zero-strength associate of TL), and two types of associative relationship between OL and TL (bidirectional and uni-

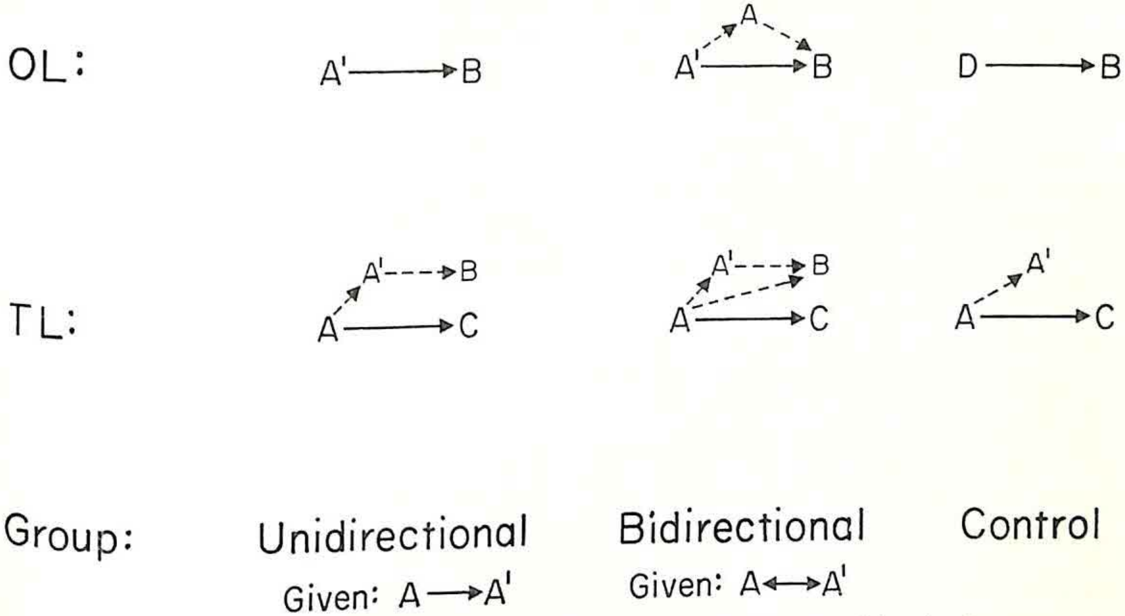


FIG. 1. Mediational processes assumed to occur in test list for unidirectional, bidirectional, and control paradigms.

directional). The term bidirectional is operationally defined as meaning that, as judged from word association norms, the words on the TL will elicit the words on the OL as associative responses and vice versa. The term unidirectional indicates that while the TL words will elicit OL words as associative responses, the reverse is not true. That is, there is no associative relationship from OL to TL.

Lists.—Both the OL and TL consisted of eight paired associates, with words as stimuli and single letters as responses. The OL was varied to meet the particular associative requirements of that group. Two different TLs were used—one for the unidirectional conditions and one for the bidirectional conditions. The associative relationship requirements of the design made it impossible for the stimulus terms of the two TLs to be identical, but the response terms were the same on both. An example from the primary unidirectional OL and TL lists is: OL, SMOKE-*F*; TL, TOBACCO-*R*, where the SMOKE → TOBACCO associative strength = .0% and the TOBACCO → SMOKE associative strength = 51.1%. An example from the comparable bidirectional lists is: OL, SWEET-*F*; TL, SOUR-*R*, where SWEET → SOUR = 43.1% and SOUR → SWEET = 56.3%.

Each of the unidirectional and bidirectional conditions had three different OLs comprising the three different levels of associative relationship between OL and TL. The mean primary strength (A → A′) for the unidirectional list was 39.12, $SD = 16.55$. For the bidirectional list, the mean primary strength (A → A′) was 39.75, $SD = 15.50$, and the reverse strength (A′ → A) was 32.50, $SD = 15.35$ (Russell & Jenkins, 1954). For the two weak-strength lists, the stimulus terms of OL were minimally related to the stimulus terms of TL (associative frequency of 2–3/1008). The same unrelated OL was used for both the unidirectional and bidirectional control conditions; there was no associative relationship between these OL stimuli and either TL. The words of the two TLs and five OLs (four experimental and one control) were equated as closely as possible for familiarity, with the majority being of AA frequency as rated by Thorndike and Lorge (1944).

Several other restrictions were applied in choosing these words. As indicated in the preceding, for the unidirectional primary- and weak-strength lists, none of the OL words elicited any of the TL words as associative responses. That is, there was an associative relationship from TL to OL, but not from OL to TL. For both unidirectional and bidirectional weak associative lists, there was no associative relationship between any OL word and any TL word, other than the one word to which the OL term was a weak associate. For both the unidirectional and bidirectional control conditions, there was no associative relationship in either direction among any of the OL words and any of the words of either TL.

Both unidirectional and bidirectional lists used the same response terms. These were single letters (consonants, including Y), with no duplications from OL to TL. Each set of eight letters had equal U-count frequencies (Underwood & Schulz, 1960) and equal distributions throughout the alphabet. None of the response letters appeared as part of the stimulus term with which it was paired. In addition, none of the letters learned successively sounded alike, nor did they form common words or abbreviations. There were two pairings of the stimulus and response terms for each list.

Procedure.—The lists were presented for learning (anticipation method) via a Stowe memory drum at a 2:1 rate, with a 3-sec. intertrial interval and 60 sec. between OL and TL. Four different orders of presentation were used, with each order used as the starting order equally often. OL learning was to one of three criteria: low (3/8), moderate (7/8), or high (two perfect successive trials: 8/8 + 8/8). TL learning was continued, after a study trial, until *S* reached a criterion of one perfect trial or for seven trials, whichever took longer.

Assignment to condition was in blocks of nine, with one *S* from each OL criterion condition and each associative strength condition per block. Over each set of four blocks, the two pairings of the unidirectional and bidirectional lists were presented equally often. The running order within blocks was determined by means of a table of random numbers, as was the assignment to starting order.

Subjects.—There were 16 *S*s in each of the 18 groups. The *S*s were students of introductory psychology at the University of California, Berkeley, who were not necessarily naive to verbal learning experiments. Twelve *S*s were discarded, either because English was not their native language or for failure to learn the OL in 30 trials.

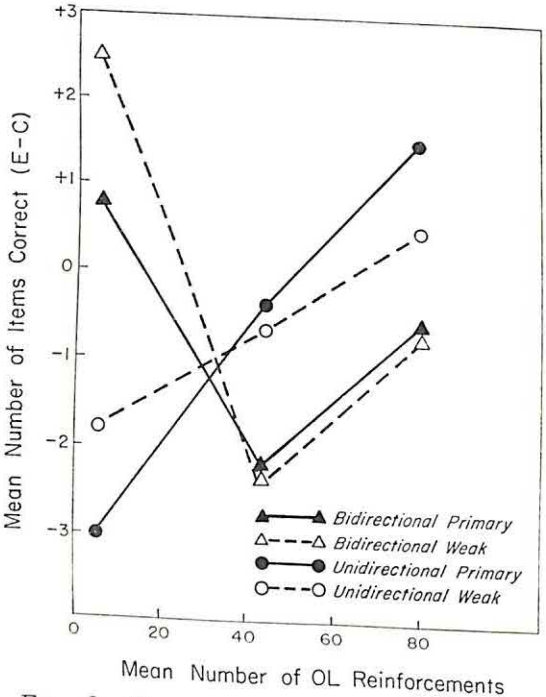


FIG. 2. Mean number of items correct on test list, Trials 1 and 2 (experimental minus control). (For explanation of bidirectional low conditions, see text.)

It should be noted that since this criterion was applied to all OL conditions, it is possible that some bias was introduced by eliminating slow learners from the high- and moderate-OL-criterion groups but not from the low OL group, since no *S*s failed to reach the low criterion in 30 trials. However, with one exception discussed subsequently, there was no evidence that the groups differed in PA learning capacity as judged from performance on the OL.

RESULTS

First-list learning.—Different first lists were used for each of the experimental conditions and for the control conditions, so that differences in list difficulty could not be ruled out. However, all groups can be compared on the number of trials required to reach the minimal criterion of 3/8. The overall mean was 3.41, with a range of means for the 18 groups of 2.50–4.62. An analysis of variance indicated that there was no significant difference between lists, F (5, 270) = 1.09. Nor were there any significant within-list differences among the three OL criterion groups, with the exception of the bidirectional control list, F (2, 270) = 5.27, $p < .01$. This difference was due to the slow learning of the low-criterion group on this list.

The 12 groups who continued learning to at least 7/8 criterion were compared for trials required to reach this criterion. The overall mean was 10.92, with a range of 8.06–14.38. There were no significant differences between lists, F (5, 180) = 1.12, or within lists, the largest F being that of the unidirectional control list, F (2, 180) = 2.11. These results support the previous assertion that it was the low-criterion group (eliminated from the present analysis) that was responsible for the significant difference in the previous bidirectional control comparison.

Second-list learning.—Performance on the TL was evaluated by determining the number of pairs correctly anticipated on the first and second test trials, by determining the number of trials to criterion, and by using several error measures. In all cases, the low-criterion bidirectional control group showed poorer learning performance (see Table 1). Since this group also was noted

TABLE 1

NUMBER OF ITEMS CORRECT (TRIALS 1 AND 2), TRIALS TO CRITERION, OVERT ERRORS, OMISSION ERRORS, AND INTERLIST INTRUSIONS FOR BIDIRECTIONAL AND UNIDIRECTIONAL GROUPS

Item	Low OL			Moderate OL			High OL		
	Primary	Weak	Control	Primary	Weak	Control	Primary	Weak	Control
Mean number of items correct (Trials 1 and 2)									
Bidirectional	5.75	7.44	4.94[a]	6.31	6.19	8.56	8.56	8.38	9.12
Unidirectional	5.44	6.62	8.44	8.69	8.38	9.06	10.44	9.38	8.88
Mean trials to criterion									
Bidirectional	12.88	10.94	14.88[a]	8.50	8.00	6.69	7.31	7.18	7.50
Unidirectional	13.12	8.44	7.56	7.69	6.64	6.31	5.12	5.62	5.44
P (errors/opportunity)									
Bidirectional	.312	.250	.303[a]	.205	.175	.127	.129	.107	.149
Unidirectional	.284	.210	.146	.122	.146	.111	.065	.092	.086
P (omission errors/opportunity)									
Bidirectional	.558	.522	.539[a]	.552	.566	.546	.527	.503	.450
Unidirectional	.602	.503	.575	.592	.583	.541	.644	.576	.632
Mean No. interlist intrusions									
Bidirectional	2.56	1.69	2.25[a]	1.81	2.38	1.44	.69	1.12	1.50
Unidirectional	1.38	1.69	1.06	1.00	1.81	1.94	.38	1.62	1.25
P (interlist intrustions/opportunity)									
Bidirectional	.186	.136	.238[a]	.190	.275	.254	.197	.194	.144
Unidirectional	.129	.172	.183	.143	.150	.287	.078	.149	.160

[a] As discussed in the text, the results of this group are discrepant throughout.

to be significantly variable on OL, it might seem appropriate to use an analysis of covariance to adjust for differences in learning ability. However, this does not turn out to be a feasible solution since the slow learners on the OL were not the slow learners on the TL. Accordingly, in the following discussion it should be kept in mind that the differences between the low-criterion bidirectional experimental groups and their control group may be overestimated.

Acquisition of test list.—The data for TL acquisition are presented in Table 1 and in Fig. 2. Transfer was measured by the number of items correct on Trials 1 and 2 since previous studies have shown that performance on early trials is most sensitive to mediated transfer effects. However, as discussed in the following, the trials-to-criterion data indicate that the same effects are still present in later TL performance.

An analysis of variance based on Trials 1 and 2 (after logarithmic transformation due to heterogeneity of variance) indicated significant main effects for both directionality, $F\ (1,\ 270) = 6.97,\ p < .01$, and for OL criterion, $F\ (2,\ 270) = 13.55,\ p < .001$. These results can be interpreted best in terms of the significant interaction effects, OL Criterion × Associative Condition, $F\ (4,\ 270) = 2.35,\ p < .06$, and Directionality × OL Criterion × Associative Condition, $F\ (4,\ 270) = 3.15,\ p < .05$. This triple interaction is due to the difference in performance between the experimental conditions (primary and weak associates) and the control condition, i.e., is due to differential transfer effects: Directionality × OL Criterion × Experimental/Control, $F\ (2,\ 270) = 6.18,\ p < .005$. In this analysis, the (C-D) control conditions are used as the learning-to-learn reference condition, against which effects due to associative relationships are to be compared. Although it is usual to include only one such control condition, the necessity of using two test lists to study the effects of unidirectionality and of bidirectionality made it necessary to include a control group for each list at each level of OL learning.[3]

---

[3] It should be pointed out that if the deviant low OL bidirectional control group is eliminated and the data from the remaining low OL control *S*s are randomly divided to form two low OL control groups of $N = 8$ each, an analysis of variance (unweighted means solution) yields only one interaction of marginal significance: OL Criterion × Experimental/Control, $F\ (2,\ 254) = 2.45,\ p < .07$.

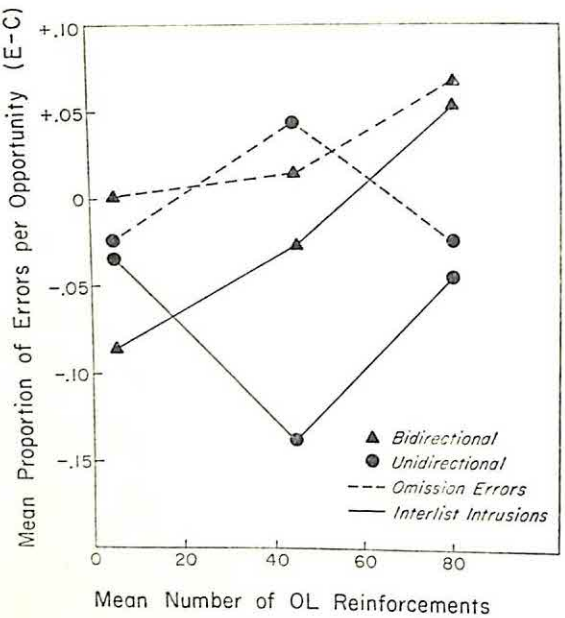


FIG. 3. Mean proportion of errors per opportunity (experimental minus control), averaged over primary and weak conditions.

As can be seen from Table 1 and Fig. 2, for the unidirectional conditions, transfer changed from negative to positive with increasing degrees of OL learning. For the bidirectional conditions, the apparent positive transfer of the low bidirectional groups (see Fig. 2) is due to the unusually poor performance of the control *S*s. In terms of the absolute number of items correct, it can be seen from Table 1 that the bidirectional and unidirectional groups are quite similar.

While the interpretation of the results from the low bidirectional groups is thus unclear, it is clear that for moderate and high degrees of OL learning, the bidirectional groups show greater negative transfer than do the unidirectional groups. Furthermore, while this negative transfer decreases with increasing OL learning, it never becomes positive.

An analysis of variance of trials to criterion (after logarithmic transformation due to heterogeneity of variance) yielded essentially the same results. Significant main effects were obtained for directionality, F (1, 270) = 14.11, $p < .001$, and for OL criterion, F (2, 270) = 32.25, $p < .001$. The triple interaction, Directionality × OL Criterion × Experimental/Control was of marginal significance, F (2, 270) = 2.95, $.05 < p < .10$.

Overt errors.—Each *S* was scored for the number of overt errors made during the first seven trials as a function of opportunity, using the formula: total errors/(56 − number of omissions), where 56 was the maximum possible correct over seven trials. This ratio provides a measure of the proportion of emitted responses that were incorrect. An analysis of variance performed on these scores (see Table 1) indicated significant main effects for directionality and for OL criterion. A higher ratio (as well as an absolutely greater number) of errors was made in the bidirectional than the unidirectional conditions, F (1, 270) = 9.14, $p < .01$. In addition, a higher ratio (as well as an absolutely greater number) of errors was made by the low-criterion *S*s than by the moderate-criterion *S*s, who in turn made more errors than the high-criterion *S*s, F (2, 270) = 22.63, $p < .001$.

Interlist intrusion errors.—The data on overt errors presented in the preceding section are comprised of several sorts of erroneous responses, i.e., intralist errors, interlist intrusions, and extralist intrusions. In the context of the mediational model being espoused in the present paper, it is clear that the interlist intrusion error is of greatest interest since this type of error provides empirical evidence for mediated interference. That is, if it is true that during the TL the mediational processes occur in the manner suggested in Fig. 1, then the presentation of A stimuli on the TL should result in the implicit elicitation of B responses, which would then interfere with the acquisition of C responses. A clear demonstration of such mediated response interference would be the *overt* occurrence of B responses during TL acquisition, i.e., would be the occurrence of interlist B intrusion errors. Hence this type of error data is of great importance to the problem presently under study.[4]

---

[4] In addition to the interlist intrusion of an IAR_2 (as B), the intrusion of an IAR_1 (i.e., A′) on the TL also would provide evidence of mediated interference consistent with the model presented in Fig. 1. In fact, such an intrusion is unlikely in the

Accordingly, each *S* was scored for the proportion of interlist intrusions making up the total overt error score (number of interlist intrusions/total overt errors). The mean values for these ratio scores and for the absolute number of such intrusions are presented in Table 1. An analysis of variance performed on the ratio scores for the 12 experimental groups indicated that there were significantly more intrusion errors for the bidirectional than unidirectional *S*s, F (1, 180) $= 22.00$, $p < .001$. The main effect of associative relationship also was significant, F (1, 180) $= 4.00$, $p < .001$, with more errors occurring when there was a weak associative relationship between TL and OL than when there was a primary relationship. Both the main effect of OL criterion and the OL Criterion × Unidirectional/Bidirectional List interaction are of marginal significance, both Fs (2, 180) $=$ 3.00, $p < .06$. The OL Criterion × Unidirectional/Bidirectional List × Associative Relationship interaction was significant, F (2, 180) $= 5.00$, $p < .01$.

The data for intrusion errors, averaged over the primary and weak association conditions (minus the control data) are plotted in Fig. 3. As can be seen, for low degrees of OL learning the unidirectional and bidirectional conditions are very similar. With increasing OL, the number of intrusions decreases for the unidirectional conditions, but increases for bidirectional conditions, indicating increasing mediated response interference with increasing A′-B strength. A Newman-Keuls comparison of the six means indicated that there was a significant difference between the unidirectional and bidirectional groups for moderate and for high degrees of OL learning, p's $< .05$; however, the difference for low OL learning was not significant.

As can be seen from Fig. 3, another measure of nonlearning (omission errors) does not show this difference between unidirectional and bidirectional groups, F (1, 180) $= 1.28$, $p > .25$. Thus it seems clear that the slower learning of the bidirectional groups can be attributed largely to mediated response interference.

DISCUSSION

The results of the present experiment support the three assumptions presented earlier, with certain limitations. Both strong (primary) and weak associates can function as IAR_1s to mediate transfer. It is less clear that the absolute amount of transfer is related to the strength of the mediator, although there is a tendency for this to be true when there is only one source of associative mediation, as in the unidirectional conditions. The presence of a second source of mediation, as in the bidirectional conditions, overshadows the effect of associative strength. Across a number of measures, for moderate and high degrees of OL learning, more negative transfer occurred in the bidirectional than in the unidirectional conditions. Since both conditions have the $A \rightarrow A' \rightarrow B$ source of interference occurring during the TL, it must be the additional $A \rightarrow B$ source of mediated interference, based on mediational IARs *that occurred during OL learning,* that accounts for the greater negative transfer in the bidirectional groups. Finally, the finding of decreasing negative transfer with increasing degrees of original learning supports the assumption that under certain conditions, *S*s can "turn off" IARs if these provide sources of interference.

The model presented in Fig. 1 is also supported by the error results. This model assumes that there will be more negative transfer in bidirectional conditions due to the additional $A \rightarrow B$ mediation of the competing response, B. The error findings, that interlist intrusions of B are greater for bidirectional than for unidirectional paradigms under those conditions in which greater bidirectional negative transfer was obtained (moderate and high degrees of OL learning), is fully consistent with the assumption. The finding that unidirectional and bidirectional paradigms do not differ in interlist intrusions when the bidirectional paradigm does not show greater negative transfer (low OL) is also consistent with this assumption, as is the finding that the two types of paradigms do not differ in omission errors, a measure that does not primarily reflect mediated interference.

---

present experiment, since IAR_1 is a word, while the correct C response is a letter, and *S*s would be unlikely to confuse items from such highly discriminable categories. Thus, the term interlist intrusion, as used in the following discussion, refers to IAR_2 intrusions only.

The prediction of greater negative transfer in the bidirectional paradigms is based on the assumption that the competing response, B, is *learned* during OL learning. Under moderate and high degrees of OL learning, it can be assumed that such learning has occurred. However, with a low OL criterion, which requires only three out of eight correct S-R pairings, it is probable that there is minimal learning of the B responses. In this case, the prediction of greater negative transfer based on OL learning would not apply. The finding that unidirectional and bidirectional paradigms do not differ in interlist intrusions of B responses under low degrees of OL learning is consistent with this notion.

The interpretation of the relative transfer effects of bidirectional and unidirectional conditions under low degrees of OL is difficult to make due to the aberrant performance of the low bidirectional control group. On the basis of the present data, the transfer effect for the low bidirectional groups can only be considered to be indeterminate. The results for the moderate- and high-OL-learning conditions are clear. With increasing degrees of OL learning, transfer changes from negative to positive for the unidirectional conditions. While the bidirectional conditions also show a decrease in negative transfer with increased OL learning, the effect remains negative even under the high OL condition.

It is suggested that increased presentation of OL stimuli increases *S*'s capacity to "turn off" their occurrence as IAR_1s during TL learning. Thus under moderate degrees of OL learning, A′-B provides no greater source of IAR_1-mediated response competition than does C-D, due to *S*'s capacity to "turn off" A′ as a response to A. With increasing degrees of OL learning, the capacity to differentiate between A and A′ is increased even more, with the result that the experimental *S* is better able to eliminate OL response competition than is the (C-D) control *S*. Thus for high degrees of OL learning, the unidirectional *S* shows positive transfer.

The *S*s in the bidirectional conditions also show this same pattern, and it is assumed that this reflects the same increasing ability to turn off the A′-B (IAR_1) source of mediated response interference with increasing degrees of OL learning. However, the second IAR_2 source of interference (A-B) remains, and hence transfer continues to be negative. The fact that high degrees of OL learning increase the number of interlist intrusion errors by strengthening the A′-B association while decreasing the overall amount of negative transfer is consistent with the findings of Mandler (1962). Another way of stating the findings regarding the unidirectional vs. bidirectional paradigms is to note that with increasing degrees of OL, *S*s are able to turn off IAR_1 sources of mediated interference, but not those due to IAR_2s. It is suggested that the difficulty in turning off IAR_2s in the present experiment was due to the fact that single-letter responses are easily confused, i.e., are poorly differentiated.

It should be noted that this finding of the importance of degree of OL learning in producing negative transfer differs from results reported by Postman (1962) and by Spence and Schulz (1965). In the latter two investigations, degree of OL learning did not differentially affect the amount of negative transfer found in an A-C, as compared to a C-D, paradigm, i.e., there was no Degree of OL Learning × Paradigm interaction. Since this interaction is found so consistently in the present study, it is important to consider possible reasons for the different results. One critical difference between the present study and those not finding a Paradigm × Degree of OL Learning effect is that in the present study an A′-B, A-C paradigm was used, while the latter studies employed an A-B, A-C paradigm. As discussed previously, in the A′-B, A-C paradigm it is assumed that the decrease in negative transfer with increased degrees of OL learning was due to *S*'s increased ability to differentiate between A and its IAR_1 (A′) during the TL. Considered from this point of view, the process occurring in the A-B, A-C paradigm is somewhat different. Here, the mediator linked to the competing response is not an IAR, but an RR. While it is tempting to think of an RR as representing the maximum point on a continuum of strength of associative relationship, which should thus produce the greatest negative effect in an interference paradigm, it may be that stimulus identity (RR) is not on the same functional continuum as IARs representing various degrees of associative relationship. From this point of view, the A-B, A-C paradigm is quite a different case than A′-B, A-C, and it would not be expected that degree of OL learning would have the same effect in the two paradigms. Just what the differing underlying mechanisms might be can only be speculated about at this point. However, considering that transfer involves list differentiation, it is apparent that to the extent that list differentiation involves differentiation of the OL and TL stimulus terms, when these differ, as in the A′-B, A-C paradigm, increased OL trials may facilitate such differentiation. However, when the stimulus terms are identical, as in the A-B, A-C paradigm, there is no need to differentiate between them, and hence increased OL trials will have no effect. Thus the need for interlist stimulus differentiation may be a critical factor in determining the effect of degree of original learning. This emphasis on *stimulus* differentiation was, of course, of major importance in Gibson's (1940) theory, and it is noteworthy that she did not include identical stimuli in her discussion of generalization.

REFERENCES

Bousfield, W. A., Whitmarsh, G. A. & Danick, J. J. Partial response identities in verbal generalization. *Psychological Reports,* 1958, **4,** 703–713.

Cramer, P. Mediated transfer via natural language associations. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 512–519.

Cramer, P., & Cofer, C. N. The role of forward and reverse association in transfer of training. *American Psychologist,* 1960, **15,** 463. (Abstract)

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review,* 1940, **47,** 196–229.

Horton, D. L., & Kjeldergaard, P. M. An experimental analysis of associative factors in mediated generalizations. *Psychological Monographs,* 1961, **75**(11, Whole No. 515).

Keppel, G. Verbal learning and memory. *Annual Review of Psychology,* 1968, **19,** 169–202.

Kimble, G. A. Mediating associations. *Journal of Experimental Psychology,* 1968, **76,** 263–266.

Kjeldergaard, P. M. Transfer and mediation in verbal learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Mandler, G. From association to structure. *Psychological Review,* 1962, **69,** 415–427.

Mandler, G., & Earhard, B. Pseudomediation: Is chaining an artifact? *Psychonomic Science,* 1965, **3,** 455–456.

Mink, W. D. Semantic generalization as related to word association. *Psychological Reports,* 1963, **12,** 59–67.

Postman, L. Transfer of training as a function of experimental paradigms and degree of first-list learning. *Journal of Verbal Learning and Verbal Behavior,* 1962, **1,** 109–118.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from Kent-Rosanoff word association test. (Tech. Rep. No. 11) Minneapolis: University of Minnesota Press, 1954.

Spence, J. T., & Shulz, R. W. Negative transfer in paired-associate learning as a function of first-list trials. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4,** 397–400.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

Underwood, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology,* 1965, **70,** 122–129.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received March 6, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 310-320

CORTICAL SLOW-WAVE AND CARDIAC RATE RESPONSES IN STIMULUS ORIENTATION AND REACTION TIME CONDITIONS[1]

WILLIAM H. CONNOR AND PETER J. LANG[2]

University of Wisconsin

The cortical slow wave and cardiac rate of human *S*s were studied in response to simple nonsignal auditory tones and to similar tones marking the onset of a reaction time (RT) foreperiod. Cortical negativity and the amplitude of the biphasic heart rate response were greatest under high-intensity and signal conditions. The initial acceleratory and secondary deceleratory components showed a positive correlation with cortical negativity. Furthermore, *S*s who reported themselves to be more alert during the task yielded higher negativity. Slow-wave activity also covaried with eyeball movement, although eyeball movement did not predict the cortical response. The slow-wave response of the brain varies systematically with cardiac rate events which have been theoretically associated both with attentional processes and with activation.

Lacey (Lacey, 1959; Lacey, Kagan, Lacey, & Moss, 1963) holds that the direction of the cardiac rate response varies with *S*'s perceptual set. Cardiac acceleration is associated with rejection of environmental input (occasioned by nocent stimuli or mental work), while cardiac deceleration accompanies sustained attention to external visual or auditory signals. In a demonstration of the latter phenomenon, Coquery and Lacey (1966) showed that the deceleratory wave which follows the presentation of a simple stimulus is further enhanced if that stimulus marks the beginning of a reaction time (RT) foreperiod. Furthermore, the point of maximum slowing just precedes the onset of the imperative event, and the degree of slowing correlates negatively with the subsequent reaction latency.

Lacey (1967) suggests that this relationship reflects an attention facilitating effect of cardiovascular feedback on cortical activity. He points out that vascular sites of central influence (e.g., the pressure sensitive carotid sinus) have been uncovered by neurophysiological research (Bonvallett, Dell, & Hiebel, 1954).

The cortical correlates of this attentional process have not been specified. However, Walter (1964, 1967) has demonstrated that slow, vertex negative waves may be recorded from the brain during a variety of conditions, in which the common element is anticipation of a response. Initially termed the expectancy wave and later renamed contingent negative variation (CNV), this slow-wave cortical phenomenon is characteristic of the RT context, with maximum negativity just before imperative stimulus onset. Furthermore, Irwin, Knott, McAdam, and Rebert (1966) and Rebert, McAdam, Knott, & Irwin (1967) found that conditions which facilitated RT produced larger CNVs.

The parallel between situations which evoke cardiac deceleration and cortical negativity encourages their evaluation in a common experimental series. The present research evaluates the hypothesis that both responses show greater amplitude after a signal stimulus than after a nonsignal stimulus with the same physical characteristics. Also, the study was designed to determine whether these responses are affected by stimulus intensity and task complexity (simple vs. disjunctive RT).

---

[1] This research was supported in part by grants to the second author from the National Institute of Mental Health (MH-10993, MH-35,324) and the Wisconsin Alumni Research Foundation. The first experiment was conducted while the first author was a National Institute of Mental Health predoctoral fellow (MH-29,611), and a portion of the data was presented in partial fulfillment of the requirements of the MS degree at the University of Wisconsin. The assistance of Ted Weerts in analyzing the data and conducting the second experiment and of Kent Hayes in developing and maintaining apparatus is gratefully acknowledged.

[2] Requests for reprints should be sent to Peter J. Lang, Department of Psychology, University of Wisconsin, Psychology Building, Charter at Johnson, Madison, Wisconsin 53706.

Even in simple nonsignal conditions, the cardiac rate response is complex, with a biphasic or triphasic arrangement of both acceleratory and deceleratory components usually observed (Lang & Hnatiow, 1962). Graham and Clifton (1966) suggest that the acceleration reflects startle or defensive reflexes, while deceleration reflects orienting behavior. Thus, a single heart rate response may include elements relevant both to "activation" and inhibitory-"attentional" processes.

In the present experiments, the significance of the obtained heart rate and slow-wave cortical waveforms was determined by trend analysis and the relationship of separate components to performance examined. Correlations between physiological systems were assessed, and the relationship between systems and subjective report of alertness was evaluated.

Two experiments are presented. The first study examined cortical and cardiac rate responses generated by auditory tones of two different intensities and by simple and disjunctive RT tasks. The second experiment replicates the first study (omitting the disjunctive RT) using a more balanced design and explores the possible role of eyeball movement as an artifact in CNV recording.

METHOD

Experiment I

The sample consisted of 24 male undergraduates who volunteered for participation in research in connection with undergraduate psychology course requirements.

Four experimental conditions were employed, each *S* receiving all conditions in one of four counterbalanced orders. In each of the two nonsignal conditions, *S*s were told to relax but remain alert while a series of 10 tones (relay closure onset) was presented. The duration of the tones was 1.92 sec., with a frequency of 800 Hz., and at approximately 80 db. for the high-intensity series and at 50 db. for the low-intensity series. In the simple RT condition, *S*s were told to respond to the second tone of each pair (signal and imperative) by pressing a microswitch as quickly as possible. Both tones were 800 Hz. and approximately 80 db. Finally, in the disjunctive RT condition, *S*s were instructed to respond to the second tone in each set only if it were the same as the first. In this latter condition, the first tone of each set was always an 800-Hz. tone of 80 db.: on 10 trials the second tone was the same, on 5 trials the second tone was 500 Hz. at 80 db., and on 5 trials it was 1,200 Hz. at 80 db. Order of presentation of second tones was independently randomized for each *S*. In both signal conditions, the second tones also lasted 1.92 sec., and the interval between offset of the first tone and onset of the second was a constant 5.76 sec. The mean intertrial interval was 58 sec., with a standard deviation of 16 sec. No intertrial intervals were less than 30 sec. The tones were produced by an Eico audio generator, amplified by a General Electric amplifier, and transmitted through a 10-in. wall-mounted Jensen speaker. Sequencing was controlled by a series of transistorized timers, and response latency was assessed by a Berkeley frequency meter.

Physiological recording.—Heart rate, EEG, respiration, and finger volume were continuously recorded throughout the experiment on a Grass Model V polygraph. EKG electrodes were placed at the lower extremities of the rib cage, and a Fels cardiotachometer converted the EKG signal to a step-function measure of rate. Silver-silver chloride Grass disc electrodes were used for the EEG signal which was recorded vertex to mastoid. The EEG signal was fed into a Grass low-level dc preamplifier which had been modified to introduce a time constant of about 8 sec. into the signal. Respiration was recorded from a strain gauge attached around *S*'s chest, and finger volume from a digital oncometer on the third finger of the left hand. The latter two variables are of only peripheral relevance to the experimental hypotheses and are not reported.

Data reduction.—The EEG and EKG heart rate signals were recorded on a Technical Measurement Corporation Model 700/1400 series magnetic tape-recording system, along with a series of pulses, each preceding stimulus onset by 3 sec. Data from the polygraph or from the recording system were transmitted to a Technical Measurement Corporation Computer of Average Transients (CAT) Model 400B for averaging over trials.

The CAT analysis provided a data point every 160 msec. along the averaged EEG and cardiotachometer curves. These points were printed out on a Hewlett-Packard automatic printer and grouped into sets of six addresses (.96 sec.) so that, for each channel, there were three sets of prestimulus information and eight sets of data up to onset of the second stimulus. For EEG data, the median of the set immediately preceding stimulus onset was defined as the base line, while for heart rate (a slower moving system), the median of the medians of the three prestimulus sets was used as the base line. Subtraction of these base lines from each of the eight poststimulus medians provided change scores which were converted, using calibration data, to beats/min or microvolts.

Procedure.—Prior to participation in the data-gathering sessions, all *S*s were given an introductory session during which an attempt was made to familiarize them with the apparatus and procedure to be used. A questionnaire was administered and discussed, the nature of the experiment explained, and *S*s were given the option of choosing not to participate (one *S* did so and was replaced). The physiological recording devices were attached, the instructions given, and four trials of each condition administered with *S* seated in a reclining chair in a darkened room. A minimum of a week later, the first of two data-gathering sessions was held. The *S* was told that part of his task throughout the session was to relax and refrain from movement. The partially soundproofed *S* room was then darkened and after a 10-min. delay for stabilization of the recordings, instructions for the first condition were given and, 3 min. later, the first trial administered. Before each subsequent condition, *E* entered the *S* room and gave the instructions for the next condition. The same procedures were used in the second recording session held at least 1 wk. later. Thus the experiment required a total of three sessions, each about 1½ hr. long.

Experiment II

The sample for the second experiment also consisted of 24 male undergraduates. A low-intensity signal condition (50 db., simple RT) replaced the disjunctive RT condition, and two channels of eye movement replaced respiration and finger volume. EEG in the second experiment was always assessed vertex to right mastoid. Silver-silver chloride electrodes attached to the forehead (vertical) and right temple (horizontal) referenced to the tip of the nose provided the eye movement channels. The signals were fed into Grass low-level dc preamplifiers which had been modified to introduce 10-sec. time constants. Change scores from these channels were calculated in microvolt units. Only one data-gathering session followed the introductory session, and all data was processed on-line by the CAT. With these exceptions, the method and procedure used in the second experiment were the same as in the first experiment.

RESULTS

Experiment I

Cardiac response curves.[3]—The heart rate response curves for the four conditions included in the first experiment are shown in Fig. 1. As for all figures in this section, the vertical axis is change from a prestimulus base, while the horizontal axis is time from stimulus onset. The first tone occurred during the first two intervals and the second tone (in signal conditions) immediately followed the last time interval score graphed.

The heart rate curves all showed an initial deceleration below base line, acceleration above base line, and secondary deceleration below base line. In nonsignal conditions, there was a return toward base line, while in signal conditions there was continued deceleration in the last interval. Trend analyses showed that the quartic component was significant in all of these curves (the range of F values with $df = 1, 20$ was 10.03–51.23, $p < .01$), accounting for 54% (50-db. stimulus) and 38% (80-db. stimulus) of the nonsignal conditions variance, but only 21% (simple RT) and 12% (disjunctive RT) of signal conditions variance. On the other hand, the quadratic trend accounted for most of the signal conditions variance—69%; simple RT, F $(1, 20) = 68.45$, $p < .01$, and 75%; disjunctive RT, F $(1, 20) = 56.25$, $p < .01$—while this quadratic component was not significant for the 50-db. nonsignal condition and accounted for only 48% of the variance in the 80-db. nonsignal condition, F $(1, 20) = 11.77$, $p < .01$. Together quadratic and quartic components accounted for more than 80% of the variance in each of the four conditions.

In an analysis of variance of these data, there was a significant conditions effect, F $(3, 60) = 4.97$, $p < .01$, and a significant Time × Condition interaction, F $(21, 420) = 7.26$, $p < .01$. By a Newman-Keuls test, only the simple and disjunctive RT conditions were not significantly different. More accelerative change occurred in the high- than in the low-intensity nonsignal condition, and in signal conditions than in nonsignal conditions. The quadratic component of the Time × Conditions interaction was significant, F $(3, 60) = 15.50$, $p < .01$, and accounted for 65% of the interaction variance. This reflects the larger quadratic component (acceleration-secondary deceleration) in signal conditions.

---

[3] An analysis of covariance design, using the prestimulus base line as a covariate was considered. However, the correlations between base line and change scores were low and inconsistent, so that such an analysis was not considered appropriate. Analyses of base-line data in both experiments parallel to those reported for change scores did not yield significant base-line effects.

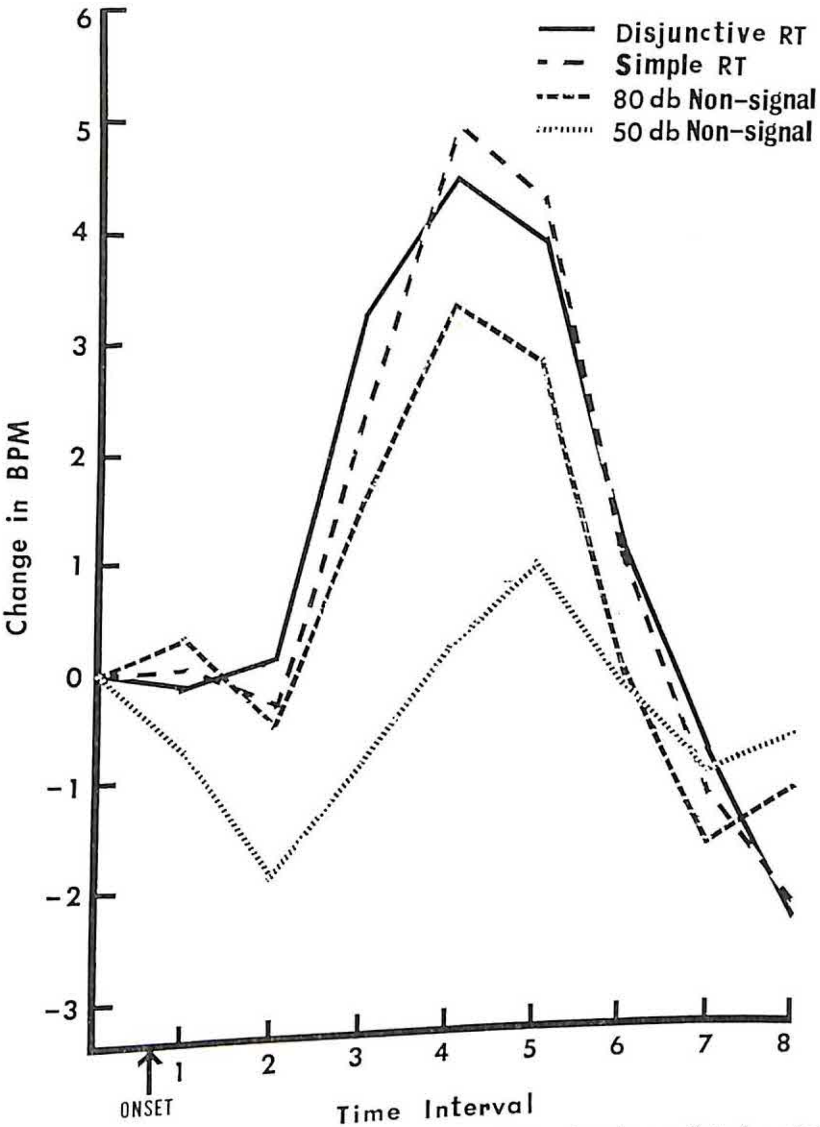


FIG. 1. Cardiac rate response curves: Exp. I. (Because of the lag in cardiotachometer recording, stimulus onset occurs approximately at the point indicated on the time scale rather than at zero.)

Cortical slow-wave response curves.—The slow-wave EEG response curves in the four conditions are shown in Fig. 2. A simple linear trend accounted for most of the variance in the 50-db. nonsignal condition (91%) and the simple RT (73%) and disjunctive RT (65%) conditions. However, this component was significant only for the low-intensity nonsignal condition, F (1, 20) = 9.17, $p < .01$. The quintic component was above or near significance ($p <$ .05) for all conditions, but accounted for only a small proportion of the variance (4–6%). Thus, there was no clearly defined wave form for the slow-wave cortical response curves.

In the analysis of variance results, there was a significant conditions main effect, F (3, 60) = 6.88, $p < .01$. By a Newman-Keuls test, there was significantly less negativity in the low-intensity nonsignal condition than in any other condition and less negativity in the high-intensity nonsignal condition than in the disjunctive RT condition. The Time × Condition interaction was also significant, F (21, 420) = 2.67, $p < .01$,

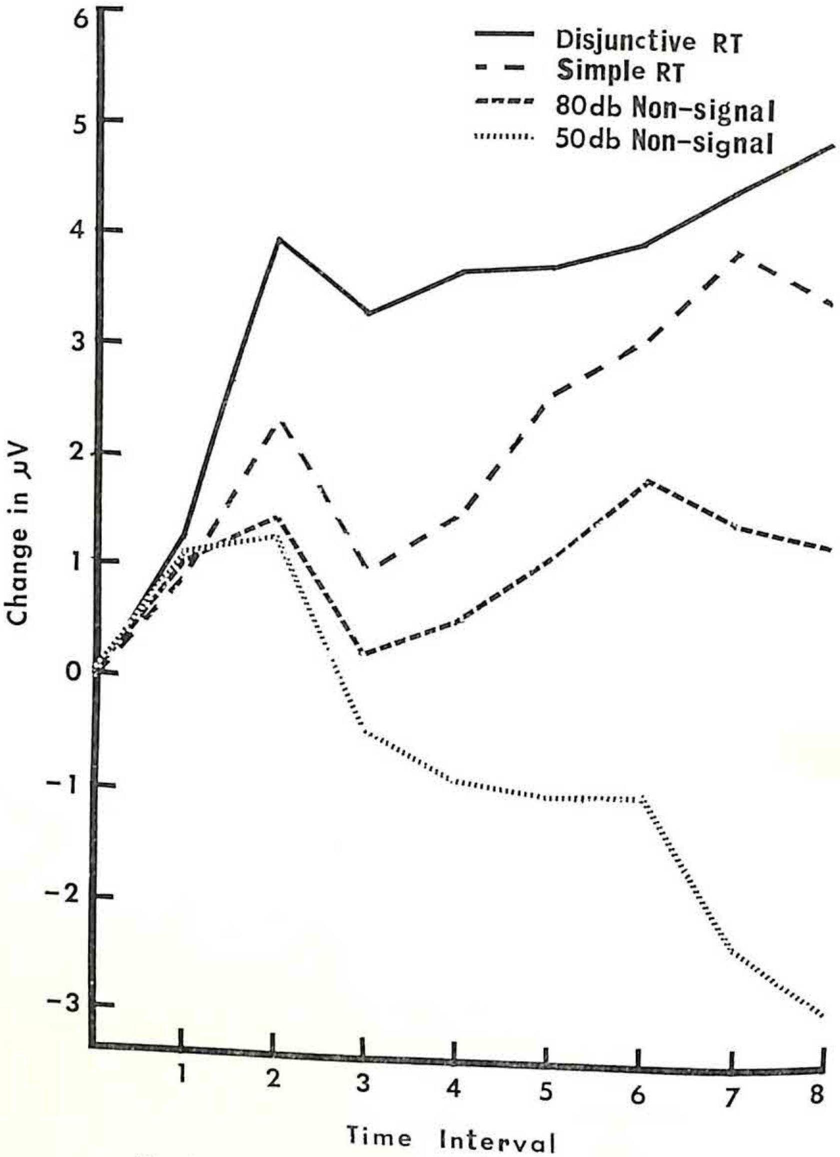


FIG. 2. Cortical slow-wave response curves: Exp. I.

with a significant linear component, F (3, 60) = 5.05, $p < .05$, which accounted for 89% of the interaction variance. The curves slope toward negativity in signal conditions and toward positivity in the low-intensity nonsignal condition.

Within-S relationships to reaction time.—For within-S investigation of the relationship between response curves and RT, the data stored on magnetic tape were averaged over five "fast" and five "slow" trials in both simple and disjunctive RT conditions for each S. In heart rate analyses, the main effect of within-S RT was significant, F (1, 12) = 5.06, $p < .05$, with more deceleration on fast trials. The main effect of within-S RT was significant for slow-wave cortical activity, F (1, 12) = 8.02, $p < .05$, and the Time × Within-S RT interaction was also significant, F (7, 84) = 2.45, $p < .05$. The slope difference between the curves was significant, F (1, 12) = 4.90, $p < .01$, and accounted for 81% of the interaction variance. Thus fast within-S RT is associated with more decelerative heart rate change and slow-wave cortical negativity.

Experiment II

The experimental design employed in the second experiment allowed for orthogonal comparison of signal (simple RT vs. nonsignal) and intensity (80 db. vs. 50 db.) effects. In addition to the cardiac and slow-wave cortical change scores, which were assessed as in the first experiment, two channels of eye movement were analyzed. For both vertical and horizontal eye movement, positive change scores reflect movement in the direction of the speaker (down and to the left of the reclining *S*).

Shape of the curves.—The shape of the cardiac response curves essentially replicated the characteristic wave form obtained in the first experiment. Quadratic, $F\ (1, 20) = 28.90$, $p < .01$, and quartic, $F\ (1, 20) = 18.52$, $p < .01$, components were again significant and together accounted for 80% of the variance. The slow-wave cortical response curves were not as similar to those obtained in the first experiment, and only the sextic component, which accounted for only 3% of the variance, was significant, $F\ (1, 20) = 4.73$, $p < .05$. Both eye movement channels showed an initial response in the direction of the speaker with recovery toward base line. The time main effect was significant for vertical eye movement, $F\ (7, 140) = 7.10$, $p < .01$, with significant linear, $F\ (1, 20) = 11.65$, $p < .01$, and quadratic, $F\ (1, 20) = 5.66$, $p < .05$, components accounting for 82% and 17% of the variance, respectively.

Signal effects.—Signal–nonsignal differences on each of the four channels of physiological information included in the second experiment were examined by trend analysis. The Time × Signal interaction was significant for the heart rate response curves, $F\ (7, 140) = 7.78$, $p < .01$, with significant linear, $F\ (1, 20) = 12.56$, $p < .01$, 61% variance, and quadratic, $F\ (1, 20) = 14.17$, $p < .01$, 31% variance, components. There was more acceleration and secondary deceleration in signal conditions. The signal main effect was significant for slow-wave cortical change scores, $F\ (1, 20) = 7.34$, $p < .05$, and there was a significant Time × Signal interaction, $F\ (7, 140) = 4.53$, $p < .01$, and linear component to the interaction, $F\ (1, 20) = 9.83$, $p < .01$, 96% variance; the curves slope toward positivity in nonsignal conditions and toward negativity in the signal conditions. In the eye movement responses curves, shown in Fig. 3, there was more eye movement in the direction of the

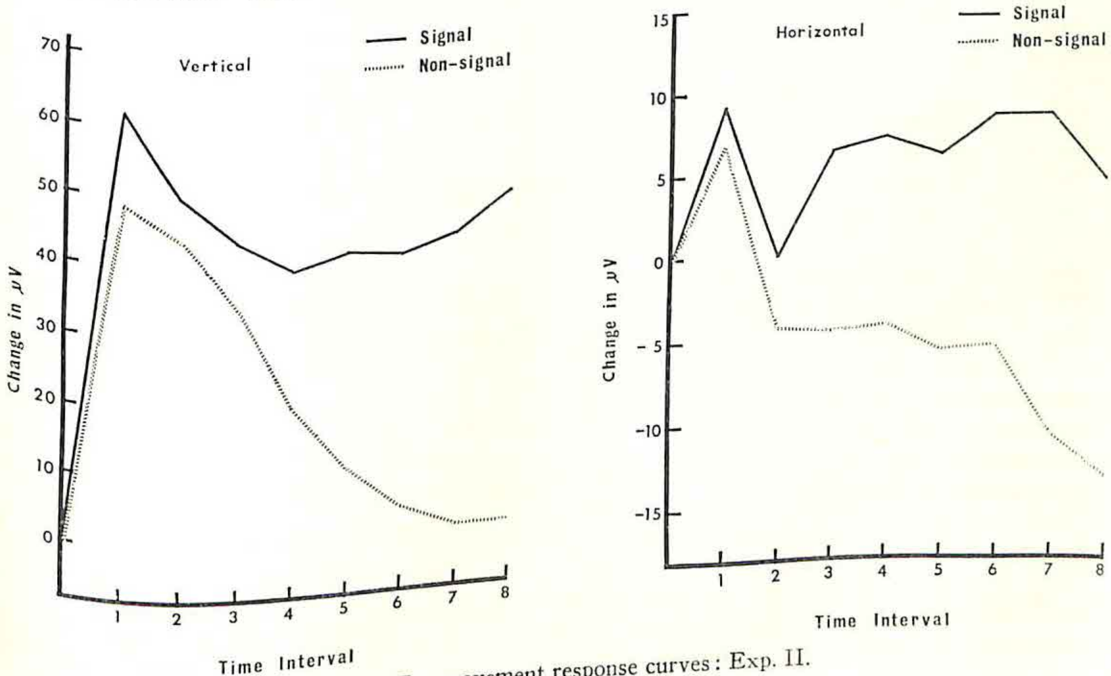


FIG. 3. Eye movement response curves: Exp. II.

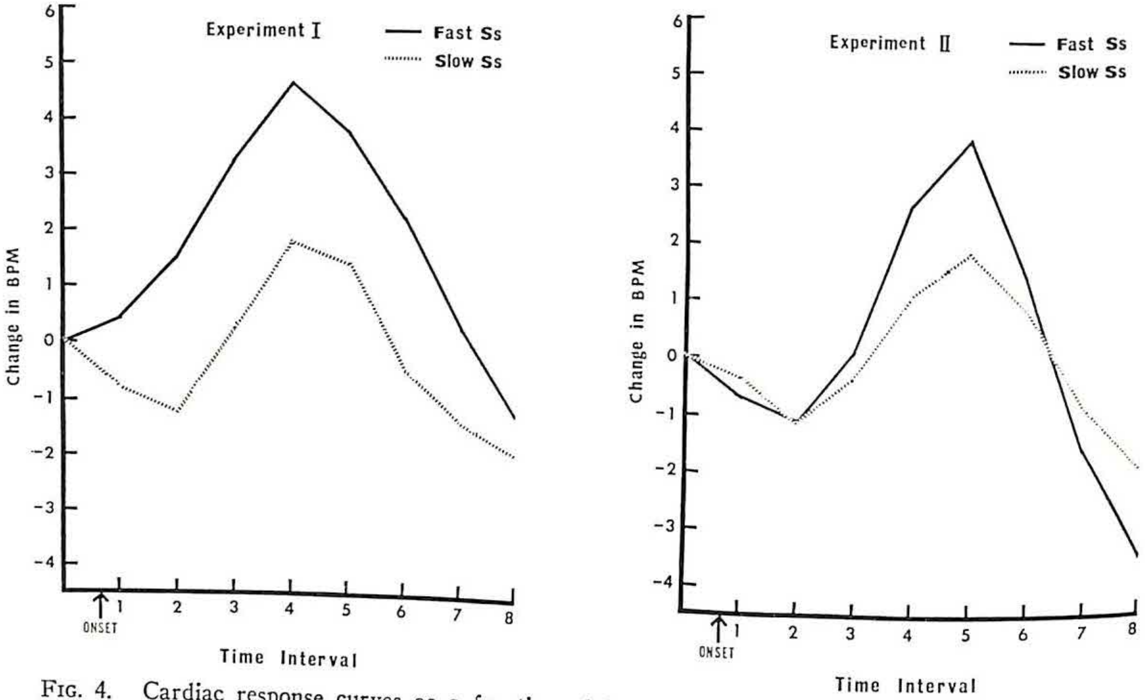


FIG. 4. Cardiac response curves as a function of between-*S* reaction time. (The *S*s in the first experiment were divided into three groups on the basis of median RT. For clarity, only the curves for the fast and slow RT groups are shown. In Exp. II, *S*s were divided into two groups on the basis of RT.)

speaker in signal conditions. For vertical eye movement, there was a significant signal effect, F (1, 20) = 4.33, $p < .05$. As the curve returned to base line in nonsignal conditions, but not in signal conditions, there was also a Time × Signal interaction, F (7, 140) = 5.10, $p < .01$; and linear component, F (1, 20) = 9.59, $p < .01$, 90% variance. The curves did not significantly differ for horizontal eye movement.

Intensity effects.—In the cardiac rate data, there was a significant Time × Intensity interaction, F (1, 20) = 4.31, $p < .01$, with linear (49% variance; $p < .05$), quadratic (21% variance; $p < .01$), and cubic (17% variance; $p < .05$) components. The curves were most disparate in the first five intervals, with more initial deceleration and less acceleration in low-intensity conditions. The main effect of intensity was significant for slow-wave cortical change scores, F (1, 20) = 5.65, $p < .05$, and the Time × Intensity interaction was also significant, F (7, 140) = 2.16, $p < .05$, with the slope difference accounting for 96% of the interaction variance. There was a slope toward negativity in high-intensity conditions and toward positivity in low-intensity conditions. Intensity was not significantly differentiated by either eye movement channel.

Between-S relationships to reaction time (Experiments I and II).—Cardiac response curves as a function of between-*S* RT (fast vs. slow *S*s) are shown in Fig. 4. In the data from the first experiment, there was a significant difference between RT groups, F (2, 21) = 5.05, $p < .05$, for the disjunctive RT condition. By a Newman-Keuls test, fast *S*s showed significantly more accelerative change than medium or slow *S*s. In the second experiment, there was a significant Time × Between-*S* RT interaction, F (7, 112) = 2.24, $p < .05$, with the quadratic component accounting for the largest portion (41%) of the variance. Thus fast responders yielded a larger accelerative component than slow *S*s in both experiments, but a larger secondary decelerative component only in the second experiment.

In both experiments, the cortical slow-wave curves suggest larger initial negativity by fast *S*s, but the curves for fast vs. slow

*S*s were not significantly differentiated in either experiment.

The eye movement response curves show fast *S*s looking down and to the right more than slow *S*s (down is in the direction of both the speaker and the response microswitch, while right is away from the speaker and toward the microswitch). For vertical eye movement, the Time × Between-*S* RT interaction was significant, $F\ (7,\ 112) = 6.85$, $p < .01$, with significant linear, $F\ (1,\ 16) = 23.86$, $p < .01$, and quartic, $F\ (1,\ 16) = 5.87$, $p < .05$, components. The curves were not significantly different for horizontal eye movement.

Level of alertness.—Degree of alertness during the experiment was assessed by a graphic scale of drowsiness anchored by "sound asleep" and "awake and alert," which *S* filled out at the conclusion of the experiment. Since the low-intensity nonsignal condition produced the most drowsiness, physiological data from this condition were used in subsequent analyses. Neither the cardiac nor the horizontal eye movement response curves of "alert" vs. "drowsy" *S*s was significantly different. For slow-wave cortical activity, there was a significant main effect of alertness, $F\ (1,\ 16) = 5.26$, $p < .05$, and Time × Alertness interaction, $F\ (7,\ 112) = 2.19$, $p < .05$, with the linear component accounting for 91% of the interaction variance. There was a slope toward greater positivity for drowsy *S*s. For vertical eye movement, the Time × Alertness interaction was significant, $F\ (7,\ 112) = 3.20$, $p < .01$, with the linear component accounting for 89% of the variance. Drowsy *S*s were looking up (i.e., away from the speaker) in the last four intervals, while alert *S*s continued to look down.

Cross-system relationships.[4]—The heart rate response clearly has three components and they were measured as follows: (*a*) *initial deceleration*—a slow valley in the first two intervals, (*b*) *acceleration*—a fast peak in the first five intervals, (*c*) *secondary deceleration*—a slow valley subsequent to the accelerative peak. For slow-wave cortical activity and eye movement, only the initial response (peak negativity within the first two intervals) was reliable and yielded significant relationships with the other variables.

In both experiments, correlations between initial cortical negativity and all three heart rate components were generally positive. However, only 5 of the 36 correlations were significant, 2 with initial deceleration (r's = .66 and .61), 2 with acceleration (.56 and .46), and 1 with secondary deceleration (.43). Cardiac activity was unrelated to either vertical or horizontal eye movement, and correlations of cortical activity to horizontal eye movement were low and inconsistent.

A clear directional relationship was found, however, between slow-wave cortical activity and vertical eye movement. The average correlation of initial response in the two systems (transformed to z, averaged, and transformed back to r) was .62 ($p < .01$), with cortical negativity associated with downward eye movement.

In order to examine this association more closely, the change scores for slow-wave cortical activity and vertical eye movement, obtained at the eight successive intervals following stimulus onset, were intercorrelated. This method yielded a matrix of 256 between-system correlations. The correlations were then divided into those between points coincident in time, slow-wave cortical "predictions" of subsequent eye movement activity, and eye movement predictions of subsequent slow-wave cortical activity. All significant correlations were positive; 14 of the 32 concurrent activity correlations (44%), 49 of 112 slow-wave cortical "predictions" (44%), and 14 of the 112 eye movement "predictions" (13%) were significant. Thus slow-wave cortical activity appears to predict subsequent eye movement activity.

---

[4] A five-page set of tables giving cross-system correlations in both experiments has been deposited with the National Auxiliary Publications Service. Order Document No. 00659 from National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10001. Remit in advance $3.00 for photocopies or $1.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

DISCUSSION

A triphasic heart rate response curve appeared in every condition of both experiments. Signal and intensity parameters significantly altered the relative amplitude of the components, but did not grossly affect their latencies. In both experiments, intensity effects had their greatest impact early in the wave form, with high intensity associated with less initial deceleration and more subsequent acceleration. As expected, signal characteristics acted primarily to increase the secondary deceleration component, although there was also more acceleration in signal conditions.

The relationship of heart rate to RT provides some support for the activation-acceleration theory (e.g., Duffy, 1962), as well as the "environmental intake" deceleration conception which prompted this experiment (Lacey, 1959; Lacey et al., 1963). In accord with activation theory, fast reactors yielded larger accelerative components than slow reactors in both experiments. On the other hand, as stimulus enhancement theory would expect, faster within-*S* RT was associated with more decelerative HR change in Exp. I (since the data in Exp. II was processed on-line, this effect could not be analyzed). Furthermore, in Exp. II, fast *S*s showed *both* more acceleration and more secondary deceleration than their slower peers. These results serve to explain Elliott's (1966) finding of both significant positive and significant negative correlations of RT with mean heart rate response. Direction of effect in data averaged over an entire foreperiod would be an arbitrary function of the wave-form component emphasized in a given trial series. However, the data reported by Coquery & Lacey (1966) are not so readily explained. They found association of fast RT with deceleration, as in the present research, but inconclusive results for acceleration. This failure to find a relationship for acceleration may stem from the rapid rate of stimulus presentation used by these investigators (every 10–20 sec.), which may have served to maintain high baseline levels of heart rate and alertness. Perhaps this masked acceleratory effects on specific trials. In any event, the present findings show the role of both components of the response, within *and* between *S*s, and provide information on their location in the cardiac wave form.

The triphasic shape of the heart rate curve does not appear to be due to a startle pattern, a possibility suggested by Graham and Clifton (1966). Detectable muscle activity (a major identifying component of the startle pattern) occurred in only a few trials; also the introductory session and the repeated measures design would be expected to minimize the likelihood of startle responses. Nevertheless, the rapid tone rise time in these experiments may have contributed, in combination with intensity, to the amplitude of the acceleratory component.

Drowsiness was considered as a potential source of artifact, since large biphasic responses are reported with sleeping *S*s (Johnson & Lubin, 1967). However, "alert" and "drowsy" *S*s (by either verbal report of EEG frequency criteria) gave comparable response curves. The lack of drowsiness effects in the present study may be due to the fact that *E* entered the room between conditions (every 10–20 min.), an interruption which Buck (1966) concludes is sufficient to restore a behavioral criterion of alertness (either detection rate or RT) to normal levels.

Although respiration data were not reported here in detail, this potential source of physiological artifact (Clynes, 1960) was examined. Significant biphasic (inspiration-expiration) respiratory curves were obtained. However, they did not reliably differentiate conditions, and the correlations between heart rate and respiratory responses were low and inconsistent. Thus, the triphasic heart rate response curve appears to be a real phenomenon of simple perceptual situations and is not readily dismissed as unique to certain states of the organism or as an artifact of another physiological system.

Slow-wave cortical response curves were extremely variable in shape, and the group curves were not characteristic of many individual records. A poor signal-to-noise ratio in this channel was apparent and the use of 10 trials per average may be near the lower limit for meaningful data. Nevertheless, amplitude of CNV significantly differentiated RT from nonsignal trials. Furthermore, *S*s produced significantly greater negativity on fast trials, although the faster *S*s could not be differentiated from slow reactors on this variable. Perhaps because of the greater difficulty in recording this variable, there was considerably more error comparing across than within individuals.

The intensity effects obtained are not consistent either with Walter's (1967) assertion that "the intensity and modality of the conditional and imperative stimuli bear no relation to the E [expectancy] wave . . . [p. 125]" nor with the finding by Rebert et al., (1967) of larger negativity responses to faint tones in both response and no-response conditions. The latter experimental result may have been due to a context effect, absent in the present study, which could override intensity determination. However, differences between the present study and previous research may stem from the fact that here slow-wave cortical responses were assessed over a longer time interval than is customary, and most of the intensity difference was found in the latter part of these curves. Intensity effects were clearly and consistently in the direction of more negativity in response to the louder tone, and this was true regardless of the intensity of the stimulus used in previous or subsequent signal trials. This result is in accord with verbal report of more altertness and attentiveness in the high-intensity conditions.

Change in the direction of positivity occurred only in the low-intensity nonsignal condition in both experiments, and the time-constant of the recording system may have been contributed artifactually to this effect. However, the amplitude of the positivity is somewhat larger than can be clearly accounted for by this factor alone. Boredom and drowsiness were heavily concentrated in the low-intensity nonsignal condition, and it is reasonable to suspect that the positivity observed in this condition may be related to fatigue. Stimuli signaled, in effect, that no new stimulus would be presented for at least 30 sec., a contingency directly opposite to the usual paradigm in which negativity is elicited.

Eye movement was found to be highly correlated with slow-wave cortical responses. However, results of an analysis of the temporal pattern were inconsistent with an eye movement artifact interpretation of the relationship. Previously recorded eye movement was relatively unrelated to subsequent cortical levels, while cortical activity predicted subsequent eye movement. Also the slow-wave cortical response curves more powerfully differentiated signal and intensity effects than did eye movement. The present findings are consistent with the interpretation that cortically-controlled eye movement responses occur reliably in the perceptual process even when the stimulus is auditory rather than visual.

The heart rate response and cortical negativity both varied with expectancy and loudness manipulations, but the direction of the effect was not consistent. The presence of an imperative stimulus elicited greater negativity and *deceleration,* while use of a high-intensity stimulus was associated with greater negativity and *acceleration.* Faster RT performance was associated with cortical negativity and both accelerative and decelerative cardiac rate components. Thus, the CNV does not clearly emerge as the cortical terminus of Lacey's (1967) feedback theory; both cardiac deceleration and acceleration relate to the vertex negative wave. If cardiac acceleration implies energy mobilization (Chase, Graham, & Graham, 1968), and deceleration a suspension of activity during attention (Obrist, 1968), cortical negativity appears to reflect both processes. It also shows sensitivity to variables not tracked by cardiac rate. Thus negativity was significantly related to subjective report of alertness and, while not significant, CNVs did tend to be larger in the more complex than in the simple RT task.

Overall, the experimental results are inconsistent with any simple, directional interpretation of the relationship between heart rate and slow-wave cortical responses. Presumably a number of relatively independent processes occur continuously, modulating both systems. Such multidimensionality is virtually inherent in the recording of slow-wave cortical activity since, as DeLucchi, Garoutte, and Aird (1962) demonstrated, the scalp acts as a spatial averager of electrical activity, transmitting all activity common to and synchronous over large areas of cortex. It is reasonable to expect multidimensionality in heart rate response as well, since a variety of factors, hemodynamic and chemical as well as neural, operate with relative independence on heart rate (Graham & Clifton, 1966). The present study highlights the complex interactions between these physiological systems and performance and suggests a need to develop current conceptualizations in the direction of a more comprehensive psychophysiological model.

REFERENCES

Bonvallet, M., Dell, P., & Hiebel, G. Tonus sympathique et activité électrique corticale. *Electroencephalography and Clinical Neurophysiology,* 1954, 6, 119–144.

Buck, L. Reaction time as a measure of perceptual vigilance. *Psychological Bulletin,* 1966, **65,** 291–304.

Chase, W. G., Graham, F. K., & Graham, D. T. Components of HR response in anticipation of reaction time and exercise tasks. *Journal of Experimental Psychology,* 1968, **76,** 642–648.

Clynes, M. Computer analysis of reflex control and organization: Respiratory sinus arrhythmia. *Science,* 1960, **131,** 300–302.

Coquery, J., & Lacey, J. I. The effect of foreperiod duration on the components of the cardiac response during the foreperiod of a reaction time experiment. Paper presented at the meeting of the Society for Psychophysiological Research, Denver, October 1966.

DeLucchi, M. R., Garoutte, B., & Aird, R. B. The scalp as an electroencephalographic averager. *Electroencephalography and Clinical Neurophysiology,* 1962, **14**, 191–196.

Duffy, E. *Activation and behavior.* New York: Wiley, 1962.

Elliott, R. Reaction time and heart rate as functions of magnitude of incentive and probability of success: A replication and extension. *Journal of Experimental Research in Personality,* 1966, **1**, 174–178.

Graham, F. K., & Clifton, R. K. Heart rate change as a component of the orienting response. *Psychological Bulletin,* 1966, **65**, 305–320.

Irwin, D. A., Knott, J. R., McAdam, D. W., & Rebert, C. S. Motivational determinants of the "contingent negative variation." *Electroencephalography and Clinical Neurophysiology,* 1966, **21**, 538–543.

Johnson, L. C., & Lubin, A. The orienting reflex during waking and sleeping. *Electroencephalography and Clinical Neurophysiology,* 1967, **22**, 11–21.

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy.* Washington: National Publishing Co., 1959.

Lacey, J. I. Somatic response patterning and stress: Some revisions of activation theory. In M. H. Appley & R. Trumbull (Eds.), *Psychological stress: Issues in research.* New York: Appleton-Century-Crofts, 1967.

Lacey, J. I., Kagan, J., Lacey, B. C., & Moss, H. A. The visceral level: Situational determinants and behavioral correlates of autonomic response patterns. In P. H. Knapp (Ed.), *Symposium on expression of the emotions in man.* New York: International Universities Press, 1963.

Lang, P. J., & Hnatiow, M. Stimulus repetition and the heart rate response. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 781–785.

Obrist, P. A. Heart rate and somatic-motor coupling during classical aversive conditioning in humans. *Journal of Experimental Psychology,* 1968, **70**, 32–42.

Rebert, C. S., McAdam, D. W., Knott, J. J., & Irwin, D. A. Slow potential change in human brain related to level of motivation. *Journal of Comparative and Physiological Psychology,* 1967, **63**, 20–23.

Walter, W. G. Slow potential waves in the human brain associated with expectancy, attention, and decision. *Archiv für Psychiatrie und Zeitschrift für die gesamte Neurologie,* 1964, **206**, 309–322.

Walter, W. G. Slow potential changes in the human brain associated with expectancy decision and intention. In W. Cobb & C. Morocutti (Eds.), The evoked potentials. *Electroencephalography and Clinical Neurophysiology,* Supplement No. 26, 1967, 123–130.

(Received March 6, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 321–326

REPETITION AND ENCODING IN SHORT-TERM MEMORY[1]

LAIRD S. CERMAK[2]

Tufts University

Three experiments were conducted to determine whether improvement on a repeated item in a short-term memory (STM) task is due to differential encoding or incremental strengthening. The Peterson distractor technique was used. It was demonstrated that although a class of repetitions tended to exhibit the same proactive inhibition (PI) effects as a new class of material, it did not allow interference from nonrepeated items to dissipate. Therefore it was concluded that initial trace strength was equivalent on each trial of the STM task and that a repetition increased the strength of that item, placing it above the others at recall.

Research in short-term memory (STM) has been increasingly concerned with the encoding mechanisms employed by *S*s. To this end, a technique has been established by Wickens, Born, & Allen (1963) that enables the researcher to determine experimentally whether or not certain types of objective classifications are also employed as subjective encoding systems.

Wickens and his associates (Wickens, 1967; Wickens et al., 1963; Wickens & Clark, 1968; Wickens, Clark, Hill, & Wittlinger, 1968) and Loess (1967) have demonstrated that performance at recall is greatly enhanced when the to-be-recalled material is from a different class than the material immediately preceding it. This has been interpreted to mean that the stimulus has been differentially encoded and thus within-class interference produced by the previously presented material is minimized.

Recent evidence (Loess & Waugh, 1967) indicates that this within-class interference dissipates over time. Furthermore, material presented during this intertrial interval does not affect the rate of dissipation as long as it is material of a different class (Nield, 1968). Any new information that does not provide further interference acts similarly to a rest interval, and the interference effects gradually diminish.

The purpose of the present set of experiments was to determine whether repeated stimuli can also allow this within-class interference to dissipate. The repeated stimuli objectively are members of the class producing the interference, yet subjectively may be differentially encoded. In order to test this possibility, three experiments were designed: first, to determine if the effects of repetition were similar to the effects of the introduction of a new class of material; second, to determine whether the effects were due simply to incremental trace strengthening; and last, to determine whether repetition acted to dissipate interference. If repetition effects were similar to the effects of the introduction of a new class, it could be concluded that subjectively a differential class had been established.

EXPERIMENT I

Method

Subjects.—The *S*s were 192 introductory psychology students at Ohio State University who chose to participate in the experiment as partial fulfillment of a course requirement. The *S*s were randomly assigned to one of the four groups in order of their appearance.

Design.—The Peterson and Peterson (1959) method was employed in which *S*s were presented a CCC trigram and then distracted from rehearsal of that trigram until recall was signaled. The next trial then consisted of a different CCC followed by the same procedure.

The stimuli consisted of consonant trigrams with association value of 29–33% (Witmer, 1935). The distractor task was counting backward by threes to the beat of a metronome (1 beat/sec). On each trial a number informed *S* where to begin counting. Neither the consonant trigrams nor the three-digit numbers contained an element that appeared in the same position in any other trigram in the list. Also, no item was repeated within a trigram. An attempt was made to equate the number of times a single consonant was employed. A counterbalancing procedure insured that each trigram appeared equally often at each trial position.

---

[1] This report is a revision of a doctoral dissertation submitted to Ohio State University. The author wishes to thank Delos D. Wickens for his assistance and use of facilities during data collection.

[2] Requests for reprints should be sent to Laird S. Cermak, Department of Psychology, Tufts University, Medford, Massachusetts 02155.

The experiment was designed to test the effects of repeating an entire block of four consecutive stimuli later in the series. Group 1 repeated Trials 1–4 on both Trials 5–8 and 9–12. Group 2 repeated Trials 5–8 during 9–12, and Group 3 repeated 1–4 during 9–12. The control group contained no repetitions. Group 2 had the same stimuli as Group 1 on Trials 5–8, and Group 3 had the same stimuli as Group 1 on Trials 1–4. All groups had the same stimuli on Trials 9–12.

Apparatus.—The *S* was seated in an IAC Model 401 sound chamber. A screen was placed before him on which the material was projected by means of a series of external mirrors. A Kodak Carousel 800 projector was timed by means of a Gerbrands tape timer, which programmed the series of events presented to *S*. The only other apparatus employed were the metronome, an Ampex stereo tape recorder, which delivered the instructions to *S*s, and four cards taped to the door of the chamber to acquaint *S* with the various types of material.

Procedure.—After the instructions had been given, *S*s practiced on the counting task until approximation to the metronome was accomplished. The door was then shut and the procedure began. The timing and sequence of events were initiated by the tape timer and followed the time intervals for one trial as follows: asterisk (ready signal), 2 sec.; trigram presentation, 1½ sec.; retention interval, 15 sec.; recall interval, 10 sec.; rest interval, 6 sec. A correct recall was scored only if all consonants were properly reproduced in the original order. Intrusions and inversions were recorded as *S* responded.

Results

The results for all four groups are depicted in Fig. 1. All within-group comparisons were analyzed by a chi square for correlated proportions (Ferguson, 1966). A chi-square technique for independent proportions (Ferguson, 1966) was used for the between-group analysis.

The critical trials were Trial 5 for Group 1 (Trials 1–4 repeated twice) and Trial 9 for both Group 2 (5–8 repeated) and Group 3 (1–4 repeated on 9–12). On Trial 5, Group 1 (77% correct) differed significantly from the control (33%), χ^2 (1) = 15.28, $p < .001$. For Trial 9 performance, Group 2 was significantly above the control (67% vs. 42%), χ^2 (1) = 5.08, $p < .05$, as was Group 3 (71%), χ^2 (1) = 7.15, $p < .01$. All points on the figure after the particular

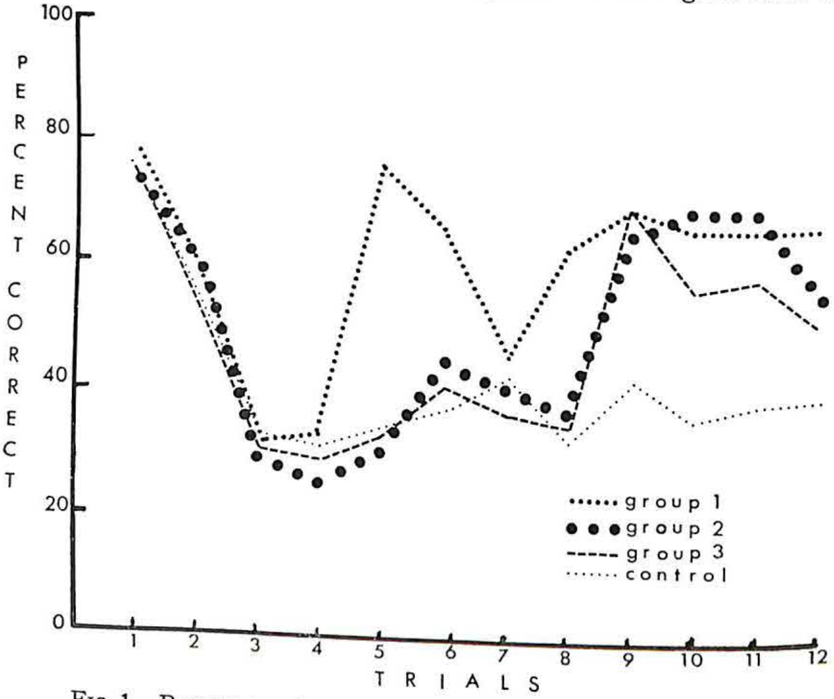


FIG. 1. Percentage of correct responses for all groups in Exp. I.

critical trial for Groups 1 and 2 were significantly above the control, at least at the $p < .01$ level, except Trial 7 for Group 1 and Trial 12 for Group 2. No trials for Group 3 were significantly better than the control other than the first repetition, the critical trial.

EXPERIMENT II

Experiment I indicated that performance was enhanced by repetition. Evidence that proactive inhibition (PI) was produced within this block of repeated items was indicated by the trend within each of the repeated blocks. The fact that the final item in one repeated block was recalled could be due to its being protected from possible retroactive inhibition (RI) produced by the design. If this RI is held constant, then it would be possible to determine whether performance on a repeated item is due to incremental trace strengthening or differential encoding. If items from different portions of the list are repeated a constant number of trials later, then the same PI curve should be regenerated at a somewhat higher level. However, if the effects are due to differential encoding, performance on a repeated item should be maximal regardless of initial presentation position. Experiment II was designed to determine which of these two alternatives was correct.

Method

Subjects.—The *S*s were 100 introductory psychology students at Ohio State University who chose to participate in the experiment as partial fulfillment of a course requirement. The *S*s were randomly assigned to one of the four groups in the order of their appearance.

Design.—The experiment was designed such that the to-be-repeated item occurred initially on either the first, second, or third trial. It was then presented twice again at four-trial intervals. Group 1 repeated the stimulus, originally shown on Trial 1, both on Trial 5 and Trial 9. Group 2 repeated the Trial 2 stimulus on 6 and 10. Group 3 repeated the Trial 3 stimulus on 7 and 11. The control group contained no repetitions. On all trials except the repetition trials all groups received the same stimuli.

Apparatus and procedure.—The apparatus and procedure were precisely the same as in Exp. I, with the exception of two additional CCCs of the same Witmer (1935) association value as before. Instructions, task, and procedure for data collection remained the same.

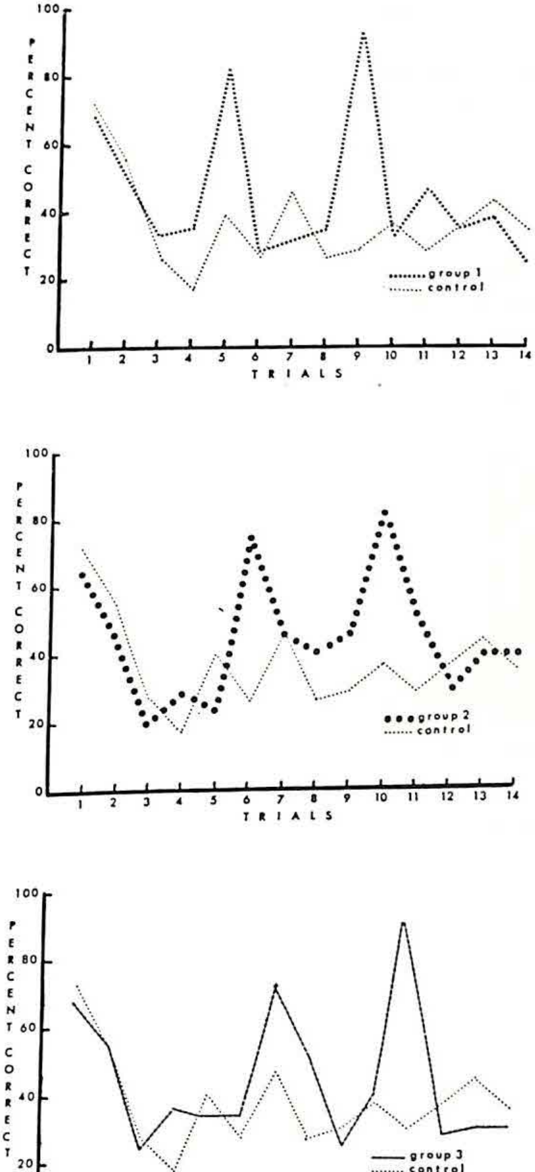


FIG. 2. Percentage of correct responses for Groups 1 (top), 2 (center), and 3 (bottom) vs. the control group in Exp. II.

Results

The results are shown in Fig. 2 (top, Trial 1 repeated on 5 and 9; center, Trial 2 repeated on 6 and 10; and bottom, Trial 3 repeated on 7 and 11). All trials on which a repeated item occurred were significantly above the control, at least at above the $p <$

.05 level. The percentage correct for the experimental groups varied from 71% to 93%, while the control varied between 25% and 43%. There were no significant differences among the experimental groups in recall of the repeated items.

EXPERIMENT III

Experiment II demonstrated that performance in recall of a repeated item is maximal regardless of initial performance level. This appeared to indicate that the effects of repetition were similar to those produced by the introduction of a new class of material and were not what would be predicted by incremental trace strengthening. However, since a class of material acts to dissipate interference generated by a previously introduced class of material, Exp. III was designed to determine whether repetition would also act in this fashion.

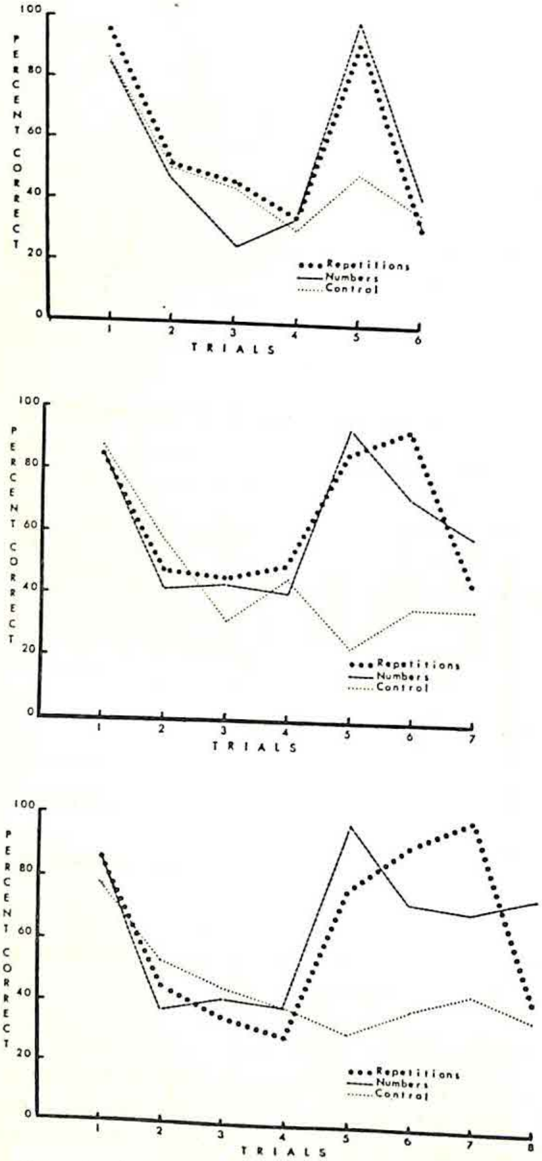


FIG. 3. Percentage of correct responses after 35 (top), 70 (center), and 105 (bottom) sec. of intervening activity.

Method

Subjects.—The *S*s were 450 introductory psychology students at Ohio State University who chose to participate in the experiment as partial fulfillment of a course requirement. The *S*s were randomly assigned to one of nine groups in order of their appearance.

Design.—Three experimental groups were employed in which the CCC occurring on Trial 1 was repeated either once, twice, or three times successively. These repetitions occurred following the fourth trial, as in Exp. II, and were followed in turn by a novel CCC. This procedure resulted in either 35, 70, or 105 sec. between presentations of nonrepeated information. To determine if the repeated item would act as a new class, three groups were designed in which either one, two, or three successive NNNs separated the two nonrepeated CCCs. The control group had one, two, or three nonrepeated CCCs between the critical test items. This resulted in nine groups of 50 *S*s each.

The distractor task employed in Exp. III was a color-naming procedure. Due to the presence of the to-be-recalled NNNs in three of the groups, a counting task could not be used. Four slides containing three rows of three colors each were projected on the screen, and *S* was instructed to call out the colors on the screen as quickly as possible. Each slide was presented for 3.75 sec. apiece; thus the retention interval remained the same as in Exp. I and II.

Apparatus and procedure.—The apparatus and procedure were essentially the same as in Exp. I and II, with the exception of the inclusion of the color slides. The metronome was eliminated, as the color-naming task was self-paced rather than timed. The instructions were altered to explain that either letters or numbers could appear as the to-be-recalled information, and the color-naming instructions replaced those of counting backward. Before data collection was initiated, *S*s were given three practice color slides to familiarize them with the task before the actual experiment began.

Two types of repeated items were employed: one-half of the *S*s received either XTN or SBM as the repeated item, and one-half received either RHJ or WLQ. This was done to determine if acoustic interference was operative in this design. XTN and SBM both contain one item from each of

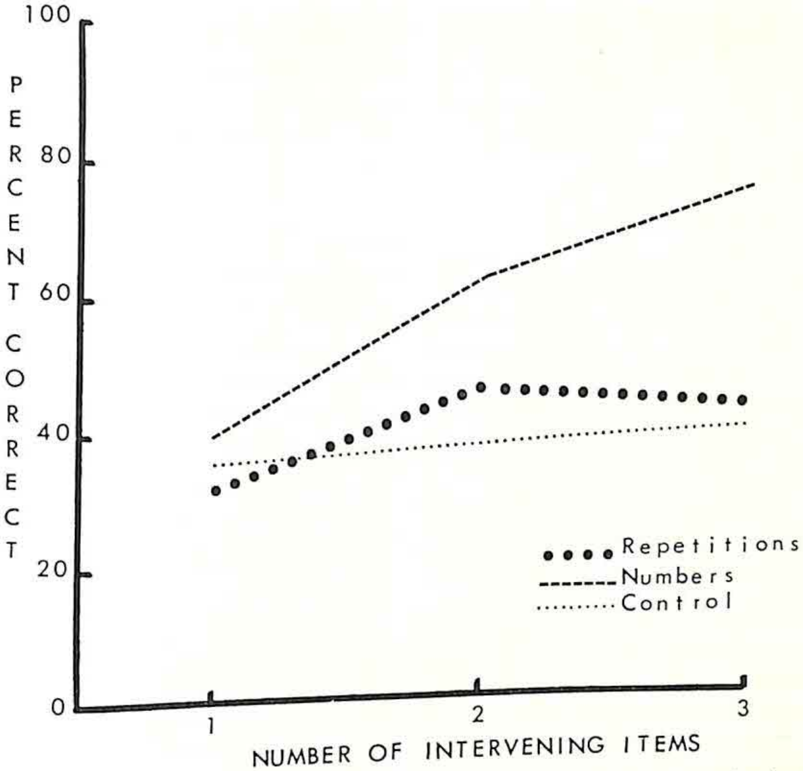


FIG. 4. Percentage of correct responses after varying numbers of intervening items.

Conrad's (1964) three classes. RHJ and WLQ contain items that are part of none of these classes. It was hypothesized that if XTN and SBM resulted in a greater maintenance of interference than did RHJ or WLQ, then the interference was acoustic. This acoustic interference would be maintained by the presentation of acoustically interfering items, even if it were the same item on each trial. If both types of items maintained interference, then interference is class-specific. If neither maintained interference, then a repeated item forms a new encoding class.

Results

Figure 3 shows the results obtained for the 35-, 70-, and 105-sec. intervals from top to bottom, respectively. This interval refers to the amount of time the intervening activity took between the presentation of the last new information and the critical test item. In all three graphs, performance is significantly lower on Trial 4 than Trial 1 at the $p < .001$ level. Also, in all graphs the effect of the intervening activity was significantly greater than that for the control group, at least at the $p < .02$ level for all trials of this activity.

Figure 4 is the critical graph for Exp. III since it is a representation of performance on the first item presented following the intervening activity. The 1, 2, and 3 on this graph indicate the number of items intervening between the last new item (on Trial 4) and this, the critical, item (Trials 6, 7, or 8). There were no significant differences among groups after one intervening item. After two presentations, the number group's performance on the CCC following the two number presentations was significantly better than that of the control (62% vs. 38%), χ^2 (1) = 4.84, $p < .01$. It was not, however, significantly better than that of the repetition group (62% vs. 46%). The repetition group was not significantly above the control either. After three intervening items, the number group was significantly above the control (76% vs. 40%), χ^2 (1) = 11.86, $p < .001$, and the repetition group (76% vs. 44%), χ^2 (1) = 9.38, $p < .01$. Again the repetition group was not significantly above the control.

The division of the effects on the critical item after repetition on either the acoustically interfering items or the nonacoustically interfering items showed no significant differences. This was consistent regardless of the number of intervening items.

DISCUSSION

The effect of repeating an item in an STM task is to facilitate recall of that repeated item. This facilitated recall is maximal regardless of initial recall performance. Therefore it would appear at first glance that a new class has been formed. However, this class does not allow dissipation of the class interference preceding it, which supports the notion that it is not a new class at all.

In order to explain this apparent contradiction, a reinterpretation of initial trace strength is needed. Recall performance is undoubtedly a poor indicant of an item's trace strength. Performance at recall is best interpreted as a function of the number of equally strong traces present. A repetition, added to any of these already equivalent traces, causes performance at recall to be enhanced equally regardless of initial item position.

The results of Exp. I, which approximate the PI effects of a new class, can also be interpreted as several equally strong traces interfering at recall. These traces are stronger than those of items presented only once, and thus the curves have been elevated. When only one item is repeated, as in Exp. II and III, the performance on that item is maximal, since it is the strongest item regardless of initial performance.

In conclusion, it can be stated that repetition acts to facilitate recall by adding to an already existent trace. The *S* need only have encoded the original stimulus to the extent that he can recognize that it has already been presented. Upon being repeated, the item gains enough incrementally to allow its strength to be above the other items' strength, and recall is maximal.

REFERENCES

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

FERGUSON, G. A. *Statistical analysis in psychology and education.* New York: McGraw-Hill, 1966.

LOESS, H. Short-term memory, word class, and sequence of items. *Journal of Experimental Psychology*, 1967, **74**, 556–561.

LOESS, H., & WAUGH, N. C. Short-term memory and intertrial interval. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 455–460.

NIELD, A. F. The effects of time and activity on the dissipation of proactive inhibition in short-term memory. Unpublished masters thesis, Ohio State University, 1968.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

WICKENS, D. D., BORN, D. G., & ALLEN, C. K. Proactive inhibition and item simliarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 440–445.

WICKENS, D. D., & CLARK, S. E. Osgood dimensions as an encoding class in short-term memory. *Journal of Experimental Psychology*, 1968, **78**, 580–584.

WICKENS, D. D., CLARK, S. E., HILL, F. A., & WITTLINGER, R. P. An investigation of grammatical class as an encoding class in short-term memory. *Journal of Experimental Psychology*, 1968, **78**, 599–604.

WITMER, L. R. The association value of three-place consonant syllables. *Journal of Genetic Psychology*, 1935, **47**, 337–360.

(Received March 8, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 327–331

SELECTIVE ATTENTION AND RATE OF DISCRIMINATION LEARNING AS A FUNCTION OF INTRADIMENSIONAL VARIABILITY[1]

ROBERT H. RITTLE[2] AND MARTIN R. BARON

Bowdoin College *Kent State University*

In Exp. I *S*s solved a discrimination learning task involving four shapes and either four, six, or eight colors. During training the two dimensions were redundant and relevant to reinforcement, enabling solution on the basis of either shape or color. The greater the number of colors involved, the less likely *S*s were to attain solution on the basis of color. Furthermore, *S*s who solved the task on the basis of the less variable (shape) dimension attained solution relatively quickly. In Exp. II intradimensional variability was manipulated prior to the start of the learning task by including four shapes and eight colors among a series of sample cards. The *S*s receiving this manipulation solved the color-relevant test task relatively slowly when compared to control *S*s, demonstrating selective attention to the less variable dimension among the (nonreinforced) sample stimuli.

Several factors have been shown to influence selective attention to particular stimulus dimensions in a discrimination learning task. If reinforcement in a previous task has been contingent on the levels of a given stimulus dimension, e.g., *S*s pay particular attention to that dimension in subsequent responding (Buss, 1956; Eimas, 1966). Other factors which induce selective attending to a stimulus dimension are a high level of discriminability between the different levels of the dimension (Archer, 1962; Sechrest & Kaas, 1965), novelty of variation along the dimension (Braley & Johnson, 1963; Guy, Van Fleet, & Bourne, 1966), or the availability of verbal labels for the dimension (Archer, 1962; Katz, 1963). The present study was conducted to determine the effect on selective attention of the number of different levels within each dimension of the stimulus set.

While the number of levels per stimulus dimension has not previously been investigated as a determinant of selective attention, its effect on the rate of learning has been reported by Battig and Bourne (1961). The demonstration by Battig and Bourne that *S*s solve a concept identification task most rapidly when the number of levels per stimulus dimension is relatively small suggested two further hypotheses concerning the effects of intradimensional variability, applicable when the manipulation involves a pair of stimulus dimensions within the same task. First, when the solution to a task can be based on either of two redundant relevant stimulus dimensions, one of which involves fewer levels than the other, solutions based on the less variable dimension should be more rapid than those based on the more variable dimension. Second, solution of a discrimination learning task involving a pair of redundant relevant stimulus dimensions should tend to be based on the less variable of the two dimensions.

EXPERIMENT I

Method

Subjects.—The *S*s were 81 female students enrolled in elementary psychology at Kent State University, who participated in partial fulfillment of a course requirement. With the deletion of 9 *S*s from the study for reasons described later, the final sample included 24 *S*s in each of three experimental conditions.

Apparatus.—Stimuli were presented through a $3\frac{1}{2} \times 5$ in. opening in a large screen separating *S* from *E*. The response manipulanda were two push-button switches, one labeled plus (+) and

---

[1] Experiment II was conducted with the support of a faculty research grant from Bowdoin College to the first author. The authors are grateful to Arthur L. Freundlich of Westbrook Junior College, Portland, Maine, for his kind cooperation in providing *S*s and research space required for Exp. II.

[2] Requests for reprints should be sent to Robert H. Rittle, Department of Psychology, Bowdoin College, Brunswick, Maine 04011.

the other minus (−), which were placed side by side in front of *S*. A white reinforcement light was situated above each of the response buttons.

The experimental stimuli, which varied along the dimensions of shape and color, were cut from colored construction paper and mounted on white, $3\frac{1}{2} \times 5$ in. file cards. The shapes included a $\frac{3}{4}$-in. square; a $\frac{3}{4} \times \frac{1}{4}$ in. rectangle; a symmetrical trapezoid with lower base $1\frac{1}{4}$ in., upper base $\frac{3}{4}$ in., and altitude $\frac{3}{4}$ in.; a rhomboid (parallelogram) of base $1\frac{1}{4}$ in., altitude $\frac{3}{4}$ in., and acute angle 72°; a vertical rectangle of base $\frac{3}{4}$ in. and height $1\frac{1}{4}$ in.; and a vertical trapezoid with a $\frac{3}{4}$-in. horizontal base, $1\frac{1}{4}$-in. vertical left side, and 1-in. vertical right side. The colors were numbers 801, 803, 805, 810, 811, 812, 818, 830, 855, and 857 of Hytone Art Paper Assortment No. 229. These colors were red, orange, yellow, light blue, violet, flesh, brown, green, gray, and black, respectively.

Experimental design and training procedure.—Intradimensional variability was manipulated in Exp. I by including either 4, 6, or 8 different levels of color in the discrimination learning task. To construct the stimuli for the 4-color condition, 2 positive and 2 negative colors were randomly selected from the 10 available colors. For the 6-color condition, 1 color was added to the positive set and 1 to the negative set. For the 8-color condition, a second color was added to the positive and the negative sets. The entire randomization process was performed twice to provide two distinct sets of stimuli for each condition. Since it was not necessary for the levels of shape to vary across experimental condition, one pair of shapes (square and parallelogram) was positive for all *S*s, while a second pair (rectangle and trapezoid) was negative for all *S*s. In constructing the stimuli, a shape which had been designated positive was always paired with a color which had been designated positive, and vice versa (i.e., the dimensions of color and shape were redundant). Thus the 4-color condition involved four positive stimuli (the combinations of two positive shapes with two positive colors) and four negative stimuli. The 6-color condition involved six positive stimuli (the combinations of two positive shapes with three positive colors) and six negative stimuli, while the 8-color condition involved eight positive and eight negative stimuli. Four separate orders of presentation were randomly determined for each set of stimuli, with the restriction that the first two stimuli of one order be different from the last two stimuli of the preceding order. Stimulus presentation began with one of the four random orders and proceeded successively through all four orders before returning to the starting point.

With two control factors (the specific colors used and the order of stimulus presentation), the design of Exp. I was a 3 × 2 × 4 factorial. The *S*s were assigned randomly to 1 of the 24 conditions, with the restriction that all cells were filled before any was repeated. Upon presentation of each stimulus, *S* pressed either the positive or the negative response button. If the correct button was pressed, the reinforcement light above that button lighted immediately, remaining lighted until the button was released. Approximately 1 sec. after the initiation of S's response, the stimulus card was removed from view; approximately 4 sec. thereafter a new stimulus card was presented. The training stimuli were presented successively to *S* until she reached a criterion of 10 consecutive correct responses.

Test procedure.—Upon reaching criterion, *S* was told that no further feedback would be given for her responses, but that she was to continue responding on the same basis. Eight test stimuli were then presented, each combining a color which had been positive during training with a shape which had been negative, or vice versa, so as to assess which of the two dimensions was dominant in determining *S*s' responses. The shapes and colors of the 4-color condition were used in constructing the test stimuli, since these levels were common to the 4-, 6-, and 8-color conditions during training. If six or more of an *S*'s eight test responses were color-determined, *S* was designated a "color responder"; if six or more of the test responses were shape-determined, *S* was designated a "shape responder." Three *S*s from the 4-color condition and one *S* from the 6-color condition were deleted because their test performance included an equal number of color-determined and shape-determined responses.

The test stimuli were preceded by two training stimuli, and interspersed with two additional training stimuli, to reveal any *S*s who altered their system of responding when reinforcement was discontinued. One *S*, from the 4-color condition, was deleted because she made more than one error on these control stimuli.

Upon completion of the test task, *S* was presented serially with the colors used in her experimental condition. Two *S*s from the 4-color condition, one from the 6-color condition, and one from the 8-color condition were discarded for failing to give an appropriate label to each color.

Results

The number of color responders vs. shape responders for each level of intradimensional variability is shown in Table 1. These results are collapsed across the two control factors, since neither of those factors affected the results. Consistent with the hypotheses of the experiment, Table 1 shows that a decreasing number of *S*s responded on the basis of color as a larger number of colors was included in the stimulus set. A chi-square test indicated that this inverse relation between the number of levels of the color dimension and solutions based on color was statistically significant, χ^2 $(2) = 16.22$, $p < .01$.

The mean number of trials required to solve the training task, as a joint function of experimental condition and the dimension used by S in attaining solution, is also shown in Table 1. Since the 8-color condition included only one color responder, the analysis of variance on these learning rates was confined to a 2×2 analysis: experimental condition (4-color vs. 6-color) vs. type of solution (color responders vs. shape responders). With the least-squares method used in estimating sums of squares (Winer, 1962), the analysis indicates that combined across experimental condition, Ss who solved the task on the basis of shape attained solution relatively rapidly, F $(1, 44) = 4.38$, $p < .05$. More importantly, the interaction between the factors of experimental condition and solution type was also significant, F $(1, 44) = 7.19$, $p < .05$. The latter result reflects the fact that solution was attained more rapidly by shape responders than by color responders within the 6-color condition, but not within the 4-color (control) condition. This result supports the hypothesis that Ss solving a task on the basis of the less variable of two stimulus dimensions attain sloution relatively quickly.

TABLE 1

NUMBER OF COLOR RESPONDERS VS. SHAPE RESPONDERS ON TASK 1, AND MEAN NUMBER OF TRIALS TO LAST ERROR DURING ACQUISITION

Item	Colors in stimulus set		
	4	6	8
Color responders			
N	14	8	1
$\bar{X}$	20.43	31.25	103.00
Shape responders			
N	10	16	23
$\bar{X}$	22.30	11.81	19.56

EXPERIMENT II

The finding that Ss tend to solve a discrimination learning task on the basis of the less variable of two redundant, relevant stimulus dimensions suggests the possibility that Ss pay selective attention to the less variable dimension. A tendency to pay attention to the less variable dimension may, e.g., exist because Ss have learned (preexperimentally) that the dimension involving fewer levels can be tested relatively quickly. Experiment I does not provide an appropriate test of the selective attention hypothesis, however, since the results can also be attributed to the fact that levels of the less variable dimension were presented relatively frequently during acquisition, thereby providing a disproportionately large number of reinforced S-R pairings for levels of that dimension. To investigate specifically the possibility that Ss pay selective attention to the less variable of two stimulus dimensions, the disproportionate number of reinforced S-R pairings for the less variable dimension was eliminated in Exp. II. This was accomplished by manipulating intradimensional variability (four shapes, four colors vs. four shapes, eight colors) in a set of sample cards which S merely looked at for a period of 3 sec. each. The subsequent test task, with color as the relevant dimension, involved two shapes and two colors which had not been used among the sample cards. It was predicted that solution of the color-relevant test task would be retarded when the sample cards included eight colors and four shapes, thereby demonstrating selective attention to the dimension which was less variable among the sample cards.

Method

Subjects.—The Ss for Exp. II were 49 female students enrolled in elementary psychology at Westbrook Junior College, Portland, Maine, who were paid $.75 each for their participation. With the deletion of 1 S from the study for reasons described subsequently, the final sample included 24 Ss in both the control and the experimental groups.

Apparatus.—Experiment II involved the presentation of a series of sample stimuli. These stimuli were geometric figures cut from colored construction paper, varying along the dimensions of shape and color. The 4 shapes used in the sample stimuli were an isosceles triangle with base 1 in. and height $\frac{3}{4}$ in.; a symmetrical trapezoid with upper base $\frac{3}{4}$ in., lower base $1\frac{1}{4}$ in., and height $\frac{3}{4}$ in.; a $\frac{3}{4}$-in. rhombus with acute angle of 72°; and a $\frac{1}{2}$-in. equilateral hexagon. The colors for the sample stimuli were chosen from the 10 colors used in Exp. I.

The test stimuli for Exp. II were geometric patterns varying along the bileveled dimensions of shape and color. The shapes for the test stimuli were a ¾-in. square, and a quadralateral with a ¾-in. vertical left side, a ¾-in. horizontal upper base, and a 1-in. horizontal lower base. The colors were selected so that no color used among *S*'s sample stimuli was included among her test stimuli. With the exception of the stimulus cards, the apparatus for Exp. II was identical to that used in Exp. I.

Experimental design and training procedure.—Following introductory instructions, *S* solved a practice problem to ensure that she understood the experimental procedure. The stimuli for the practice problem were 3½ × 5 in. file cards containing either one or two black dots, ½ in. from either the upper or the lower edge of the card. For all *S*s the position of the dots was relevant to reinforcement, while the number of dots was irrelevant. The procedure for administering the practice problem was the same as that used in the discrimination learning task of Exp. I. One *S*, who failed to reach criterion on the practice problem in 56 trials, was deleted from the study.

After *S* attained criterion, she was shown a set of eight sample cards which were presented successively for 3-sec. durations. To form the sample stimuli for the experimental condition, each of eight colors was combined with one of four shapes (each shape being used twice) to produce a set of eight cards. To form the sample stimuli for the control condition, half of the original eight colors were eliminated, each of the remaining colors being paired with a second shape to form the set of eight cards. Thus the sample cards used for the two conditions were comparable except that the experimental set included eight colors with four shapes, while the control set included four colors with four shapes. To balance partially any idiosyncratic effects of the particular colors used in the sample stimuli, the process described previously was repeated to form a second set of sample cards for both the experimental and control conditions. The design of the experiment, then, involved one bileveled experimental factor and one bileveled control factor. The *S*s were assigned randomly to one of the four resulting conditions, with the restriction that all conditions were used before any was repeated.

The following instructions were read to *S* prior to presentation of the sample stimuli:

> All right; the next task will be similar, but it will involve a new set of cards. Before we begin this task, I am going to show you some of the cards so you will know what they are like. Do not press the buttons when we first begin—just look at the cards carefully. Then, when I do tell you to begin pressing the buttons, try to solve the task as quickly as you can. O. K.?

Test procedure.—After presenting the eight sample stimuli, *E* informed *S* that she was to begin responding, and immediately administered the test task with the same procedure as was used for the learning task of Exp. I. The stimuli for the test task involved the combinations of two colors with two shapes, color being relevant to reinforcement and shape irrelevant. After *S* reached a criterion of 10 consecutive correct responses on the test task, she was administered a color identification task which involved the 10 colors used in the experiment. All *S*s gave an appropriate and distinctive response to each of the colors.

Results

While the major purpose of the practice task was to familiarize *S*s with the procedure of the experiment, the results of the practice task were also used to compare the learning ability of the experimental and the control groups prior to the introduction of the experimental manipulation. Contrary to expectation, the mean number of trials to last error on the practice problem was lower (6.12 vs. 11.37) for *S*s in the experimental condition than for *S*s in the control condition, $t\ (46) = 2.48$, $.01 < p < .05$, two-tailed test. However, since the superior learning ability among *S*s in the experimental group would, if anything, counteract the anticipated effects of the experimental manipulation, the test results were analyzed as planned.

On the test task of Exp. II a total of three *S*s, all in the experimental condition, failed to begin a criterion run by Trial 24. To make the variance of the two sets of scores more nearly homogeneous, the test results of these three *S*s were scored for a maximum of 24 trials. Even with this restriction, the mean number of trials to last error on the test task was 9.54 for *S*s in the experimental group, and 6.54 for *S*s in the control group, $t\ (46) = 1.69$, $p < .05$. Similarly, when the test data were analyzed in terms of the mean number of informed errors (i.e., excluding errors occurring on Trial 1), experimental *S*s committed an average of 5.12 informed errors, while control *S*s committed an average of 3.46 informed errors, $t\ (46) = 1.70$, $p < .05$. The relative difficulty of the experimental *S*s in solving the color-relevant test task apparently resulted from their attending selectively to shape, which was the less variable dimension among their sample stimuli.

GENERAL DISCUSSION

In addition to demonstrating selective attention to the less variable of two stimulus dimensions, the results of Exp. II are relevant to the theoretical distinction between associational vs. hypothesis-testing models of discrimination learning (see Bourne, 1968). According to associational theories, discrimination learning proceeds on the basis of an S-R reinforcement process; selective attention is typically incorporated into such theories on the assumption that *S*s become more likely to respond to a stimulus dimension as the levels of that dimension develop differential S-R bonds. In Exp. II, however, *S*s attended to the less variable dimension as a result of viewing the (nonreinforced) sample stimuli. Consequently, selective attention to the less variable dimension cannot, in Exp. II, be attributed to an increase in the strength of the S-R bonds for the levels of that dimension. Alternately, an hypothesis-testing model of the type described by Trabasso and Bower (1968) is more compatible with the findings of Exp. II. According to an hypothesis-testing model, *S* actively selects hypotheses to be tested in a discrimination learning situation. Prior knowledge regarding intradimensional variability might be expected to induce selective attending to the less variable dimension, according to such models, since relatively few hypotheses must be tested as a stimulus dimension involves fewer levels. Thus the findings of Exp. II appear to be more consistent with an hypothesis-testing model of discrimination learning than with current associational models.

REFERENCES

ARCHER, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. *Journal of Experimental Psychology*, 1962, 63, 616–620.

BATTIG, W. F., & BOURNE, L. E. Concept identification as a function of intra- and interdimensional variability. *Journal of Experimental Psychology*, 1961, 61, 329–333.

BOURNE, L. E., JR. Concept attainment. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

BRALEY, L. S., & JOHNSON, D. M. Novelty effects in cue acquisition and utilization. *Journal of Experimental Psychology*, 1963, 66, 421–422.

BUSS, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *Journal of Experimental Psychology*, 1956, 52, 162–166.

EIMAS, P. D. Effects of overtraining and age on intradimensional and extradimensional shifts in children. *Journal of Experimental Child Psychology*, 1966, 3, 348–355.

GUY, D. E., VAN FLEET, F. M., & BOURNE, L. E. Effects of adding a stimulus dimension prior to a nonreversal shift. *Journal of Experimental Psychology*, 1966, 72, 161–168.

KATZ, P. A. Effects of labels on children's perception and discrimination learning. *Journal of Experimental Psychology*, 1963, 66, 423–428.

SECHREST, L., & KAAS, J. S. Concept difficulty as a function of stimulus similarity. *Journal of Educational Psychology*, 1965, 56, 327–333.

TRABASSO, T., & BOWER, G. H. *Attention in learning: Theory and research*. New York: Wiley, 1968.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received March 10, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 332–338

SENTENCE PROCESSING ASSESSED THROUGH INTRASENTENCE WORD ASSOCIATIONS[1]

ROBERT W. WEISBERG[2]

Temple University

Two experiments were conducted to study sentence structure through intrasentence word association. The *S*s learned a sentence and were then given individual words from the sentence as stimuli for word association, with the other words in the sentence as possible responses. Generally, the most frequent response to each word in a sentence was a word closely related to the stimulus word in the deep structure of the sentence rather than a word in close physical proximity to the stimulus. In Exp. I, a string of English words, when given two different structures through the use of appropriate context, gave two different patterns of word associations, and these patterns corresponded to the differences in underlying structure. In Exp. II, sentences with very different surface structures but the same underlying structure gave very similar patterns of intrasentence word associations. The results from the two experiments indicated that the deep structure of sentences was the most important factor in determining the organization that a speaker gives to a sentence.

Recent research indicates that the first step in the comprehension of a sentence by a native speaker of English consists of the speaker's determination of the surface structure of the sentence. Johnson (1966) found that errors occurring when *S*s attempted to learn sentences indicated that *S*s learned the sentences not as wholes but as groups of words, which were then put together to form the entire sentence. These groups of words corresponded to the smallest constituents in the surface structure. Garrett, Bever, and Fodor (1966), using a psychophysical technique, also found that *S*s broke sentences into units that seemed to correspond closely to the major constituents in the surface structure. Similar results have also been found by Ammon and his associates (Ammon, 1968; Suci, Ammon, & Gamlin, 1967).

However, other workers, when studying long-term storage and comprehension of sentences, have found that sentences seem to be stored in terms of their deep structures rather than surface structures. Sachs (1967) found that changes in the surface structure of a sentence were not noticed by *S*s when they tried to recall whether a given sentence was one which they had heard earlier. Sachs felt that these surface structure changes were not noticed because *S*s stored the sentences on the basis of deep structure and so forgot the surface structure. Blumenthal (1967) and Rohrman (1968) have also found evidence that sentences are stored on the basis of their deep structures.

One tentative conclusion that can be drawn from these various studies is that a speaker, on hearing a sentence, first determines the surface structure of the sentence. From this the speaker constructs the deep structure, which then serves as the basis for the interpretation and subsequent storage of the sentence. The present paper is a report of two experiments that investigated the psychological organization of sentences through the use of intrasentence word associations. This organization was then compared with the deep structures of the sen-

---

[1] This study is based partially on a PhD thesis submitted to the Department of Psychology of Princeton University. Experiment II was supported in part by United States Public Health Service Grants MH-10742-03 and HD-01910-02 to Princeton University (S. Glucksberg, principal investigator). Parts of this paper were presented at the meeting of the Eastern Psychological Association, Boston, April 1967. The author is indebted to S. Glucksberg for his aid and advice. Thanks are also due to Jerry M. Suls, who conducted Exp. I., which was supported in part by Temple University Research Grant 400-97-68 (R. Weisberg, principal investigator).

[2] Requests for reprints should be sent to Robert W. Weisberg, Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.

TABLE 1

Sentences Used in Exp. I

Sentence number	Sentence	Predicted intrasentence word associations
1	Places with new planes flying can be dangerous.	FLYING → *planes* DANGEROUS → *places*
2	Even with new planes, flying can be dangerous.	FLYING → *dangerous* DANGEROUS → *flying*
3	The man sent to train George drove furiously to the station.	TRAIN → *George* GEORGE → *train* DROVE → *station*
4	In order to catch his train, George drove furiously to the station.	TRAIN → *catch* GEORGE → *drove* DROVE → *George*
5	Your hope of marrying Anna was surely impractical.	MARRYING → *Anna* ANNA → *marrying* IMPRACTICAL → *hope*
6	In her hope of marrying, Anna was surely impractical.	MARRYING → *hope* ANNA → *impractical* IMPRACTICAL → *Anna*

tences as postulated by transformational grammarians (Chomsky, 1965). In both experiments, *S*s were presented with a novel sentence that they were asked to learn by writing the sentence a number of times from memory. Words from the sentence were then presented individually as word association stimuli, and *S*s were asked to respond with one of the other words from the sentence, whichever word came to mind. Both experiments were concerned with the general hypothesis that the most frequent response to each stimulus word in word association would be the word most closely related to it in the deep structure of the sentence.

Experiment I

The sentences used in Exp. I are presented in Table 1. These sentences were three pairs taken from those used by Garrett et al. (1966). Each pair of sentences is made up of the same string of words, except for the first one or two words. However, the two members of each pair have very different deep structures. These deep structures can be represented as sets of propositions as follows:

Sentence 1. Places are dangerous. Planes are new. Planes are flying.
Sentence 2. Planes are new. Flying is dangerous.
Sentence 3. The man was sent. The man drove to the station. The man trained George.
Sentence 4. George caught his train. George drove to the station.
Sentence 5. You hoped. You married Anna. You were impractical.
Sentence 6. Anna hoped. Anna married. Anna was impractical.

These propositions are combined by means of transformational rules in order to obtain the actual sentences in Table 1. On the basis of these deep structures, it would be predicted that the patterns of intrasentence word associations would differ for the members of each pair of sentences. The predicted associations are also presented in Table 1.

Method

Subjects.—Thirty-five Temple University undergraduates, divided into a group of 17 and a group of 18, served as *S*s. All *S*s were native speakers of English and were experimentally naive. All *S*s were unpaid volunteers. Approximately the same number of males and females were included in each group.

Procedure.—Sentences 1, 3, and 5 were presented to one group, and the other sentences were presented to the second group. Each *S* was tested individually. The *S*s were scheduled at their convenience and were randomly assigned to groups according to their order of appearance at the laboratory. The *S* was instructed that he was to learn "some sentences." The sentences were learned one at a time. The *S* read the sentence from a card provided by *E* and then returned the card to *E*. The *S* then wrote the sentence from memory. The *E* indicated to *S* whether the written response was a correct reproduction of the sentence. The *S* was then given the card again, asked to read the sentence aloud, etc. This procedure was continued until *S* wrote the sentence correctly from memory five consecutive times.

After *S* had learned the first sentence to criterion, he was given the word association test for that sentence. Not all the words from each sentence were used as stimuli or responses in the word association. The words used as stimuli are those presented in the right-hand portion of Table 1. In order to reduce response variability, some of the words were underlined when *S* was shown the sentence. (These words are underlined in the sentences in Table 1.) The *S*s were instructed to write the sentence from memory with those words underlined, and the underlined words were the only words that were acceptable responses during word association. That is, immediately preceding the first word association test, *S* was told to respond to each stimulus word with one of the underlined words, the first one of these that came to mind. In two pairs of sentences the identical words were underlined, while in the third pair three of the words were the same, and the fourth words were in corresponding physical positions. The *S*s had no knowledge as to which words would be used as stimuli in the word association test.

The procedure can be summarized as follows. The *S* read the first sentence and wrote it from memory (including the underlining) to a criterion of five correct reproductions. The word association test for that sentence was then given. This was repeated twice more for each *S*. The order of the sentences presented to each *S* and the order of the stimuli for the word association were all randomized for each *S*.

Results and Discussion

Since the sentences were presented individually to *S*s and since the sentences were neither excessively long nor excessively complicated, essentially no errors were made during learning by any *S*. Consequently, no analysis of errors during learning could be made. The results of the intrasentence word association are presented in Table 2. In each case the frequency of responses is compared for the same stimulus word in the two members of each pair of sentences. The responses of interest are the predicted responses for each of the separate words. As can be seen in Table 2, the results were generally as predicted. For all eight of the stimuli, the most frequent response was the underlined word most closely related to the stimulus word in the deep structure of the sentence. Only the responses to MARRYING in Sentences 5 and 6 are not significantly different.

One further point of interest concerns the relationship between actual physical separation of stimulus and predicted response and the frequency of response obtained. On the basis of research in verbal learning, it would be predicted that as the physical distance (the number of intervening words) increases between stimulus and response, the associative connection should become weaker. This would result in a decreased frequency of association between the more distant words in the sentences. On the basis of the deep structures of the sentences, no such prediction would be made since physical distance is of little importance in determining sentence structure. In order to look at the relationship between physical distance and frequency of association, the distance between each stimulus and its predicted response was determined by counting the number of intervening words. A rank-order correlation was then computed between these distances in the respective frequencies of response. This correlation was +.04, which indicates that the physically closer responses were not necessarily the most frequent.

These results provide some support for the hypothesis that the deep structure of a sentence determines the organization given to that sentence, at least to the extent that sentence organization or structure can be assessed through intrasentence word association. Identical groups of words, when given different deep structures through appropriate context, gave significantly different patterns of associations, and these associations corresponded to predictions made on the basis of the deep structures of the sentences.

EXPERIMENT II

The results of Exp. I, while in principle supporting the deep structure hypothesis, are somewhat ambiguous. In the case of at least some and perhaps all of the word association stimuli, the same response would be predicted on the basis of the surface struc-

TABLE 2

Intrasentence Word Associations Obtained in Exp. I

Responses to FLYING		
Sentence 1.	*Places* with new $\overset{10}{\textit{planes}}$ *flying* can be $\overset{2}{\textit{dangerous}}$.	$p < .05$[a]
Sentence 2.	*Even* with new $\overset{6}{\textit{planes}}$, *flying* can be $\overset{11}{\textit{dangerous}}$.	
Responses to DANGEROUS		
Sentence 1.	$\overset{8}{\textit{Places}}$ with new *planes* $\overset{0}{\textit{flying}}$ can be *dangerous*.	
Sentence 2.	$\overset{2}{\textit{Even}}$ with new *planes*, $\overset{9}{\textit{flying}}$ can be *dangerous*.	$p = .005$[b]
Responses to TRAIN		
Sentence 3.	The man $\overset{3}{\textit{sent}}$ to *train* $\overset{7}{\textit{George}}$ *drove* . . . to the *station*.	
Sentence 4.	In order to $\overset{9}{\textit{catch}}$ his *train*, $\overset{2}{\textit{George}}$ *drove* . . . to the *station*.	$p = .025$[b]
Responses to GEORGE		
Sentence 3.	The man *sent* to $\overset{5}{\textit{train}}$ *George* $\overset{7}{\textit{drove}}$ . . . to the *station*.	$p < .025$[a]
Sentence 4.	In order to *catch* his $\overset{0}{\textit{train}}$, *George* $\overset{16}{\textit{drove}}$ . . . to the *station*.	
Responses to DROVE		
Sentence 3.	The man *sent* to *train* $\overset{2}{\textit{George}}$ *drove* . . . to the $\overset{11}{\textit{station}}$.	$p = .025$[b]
Sentence 4.	In order to *catch* his *train*, $\overset{9}{\textit{George}}$ *drove* . . . to the $\overset{6}{\textit{station}}$.	
Responses to MARRYING		
Sentence 5.	Your $\overset{3}{\textit{hope}}$ of *marrying* $\overset{13}{\textit{Anna}}$ was surely *impractical*.	$p > .05$[a]
Sentence 6.	In her $\overset{6}{\textit{hope}}$ of *marrying*, $\overset{8}{\textit{Anna}}$ was surely *impractical*.	
Responses to ANNA		
Sentence 5.	Your *hope* of $\overset{12}{\textit{marrying}}$ *Anna* was surely $\overset{2}{\textit{impractical}}$.	$p < .01$[a]
Sentence 6.	In her *hope* of $\overset{5}{\textit{marrying}}$, *Anna* was surely $\overset{11}{\textit{impractical}}$.	
Responses to IMPRACTICAL		
Sentence 5.	Your $\overset{12}{\textit{hope}}$ of *marrying* $\overset{1}{\textit{Anna}}$ was surely *impractical*.	$p = .01$[b]
Sentence 6.	In her $\overset{5}{\textit{hope}}$ of *marrying*, $\overset{8}{\textit{Anna}}$ was surely *impractical*.	

Note.—Entries are the number of *S*s out of 18 (Sentences 1, 3, and 5) or out of 17 (Sentences 2, 4, and 6) who gave the indicated response upon presentation of the particular stimulus word.

[a] By chi square with 1 *df* and Yates' correction.

[b] By Fisher exact probability test.

ture of the sentences. Experiment II was conducted in order to eliminate this ambiguity in the results of Exp. I. In Exp. II, groups of sentences were used in which the deep structure was held constant, while surface structure was allowed to vary across groups. This should make it possible to specify more precisely which aspect of sentence structure was primary in determining sentence organization.

The stimuli used in Exp. II are shown in Table 3. The surface structures are different

TABLE 3

SENTENCES USED IN EXP. II

Group	Sentences
A	SLOW CHILDREN EAT COLD BREAD. KNOWN AUTHORS PREPARE CORRECT LETTERS. HONEST KINGS UNITE LOYAL DISTRICTS.
Par	SLOW CHILDREN EAT, AS YOU KNOW, COLD BREAD. KNOWN AUTHORS PREPARE, IT IS SAID, CORRECT LETTERS. HONEST KINGS UNITE, I AM TOLD, LOYAL DISTRICTS.
Pass	COLD BREAD IS EATEN BY SLOW CHILDREN. CORRECT LETTERS ARE PREPARED BY KNOWN AUTHORS. LOYAL DISTRICTS ARE UNITED BY HONEST KINGS.
Rel	CHILDREN WHO ARE SLOW EAT BREAD THAT IS COLD. AUTHORS WHO ARE KNOWN PREPARE LETTERS THAT ARE CORRECT. KINGS WHO ARE HONEST UNITE DISTRICTS THAT ARE LOYAL.

Deep structures of sentences in Exp. II

Children eat bread. Children are slow. Bread is cold.
Authors prepare letters. Authors are known. Letters are correct.
Kings unite districts. Kings are honest. Districts are loyal.

for each group of sentences, while the deep structures are essentially identical for all groups. (The deep structures of these sentences are outlined in the lower portion of Table 3.) Therefore, on the basis of the surface structures, the different groups should give different patterns of associations, while on the basis of the deep structure the pattern should be the same for all the groups. As an example, in Group pass, IS EATEN should give *bread* as a response on the basis of the deep structure and *by* as a response on the basis of surface structure.

Method

Subjects.—Eighty Princeton University undergraduates, randomly divided into four groups of 20 each, served as *S*s. None had served in verbal learning experiments before; all were paid volunteers and were native speakers of English.

Materials and procedure.—The groups of sentences are shown in Table 3. All the sentences were constructed from a set of words used by Cofer and Bruce (1965). These words were used because no word in this set is a response to any other word in the set in free association. Group A learned three simple active sentences, one at a time, and all the words from each sentence were used as stimuli for the subsequent word association tests. Group par learned the same simple active sentences as Group A, but inserted into each sentence was a parenthetical remark. Group par was given the same words as Group A during the word association test; the words from the parenthetical remarks were not used as stimuli. Group pass learned the passive sentences corresponding to the active sentences learned by Group A. There was a slight change in the stimuli presented to this group in the word association tests. Group pass was given "is + participle" as the verb (i.e., IS EATEN, etc.). The word "by" was not used as a stimulus in the word association. Finally, Group rel learned sentences in which the adjectives were replaced by relative clauses. This group was given the same words as Group A in the word association tests.

The procedure was the same as that for Exp. I. The *S*s were randomly assigned to groups according to their order of appearance at the laboratory, and order of sentence presentation and word association stimuli were randomized for all *S*s. The *S* learned the first sentence and was then given the word association test for that sentence. This pro-

TABLE 4

INTRASENTENCE WORD ASSOCIATIONS FOR GROUPS A, PAR, PASS, AND REL IN EXP. II

Stimulus	Group	Responses		χ^2	p
		children	other		
SLOW	A	19	1	4.23	>.05
	Par	19	1		
	Pass	19	1		
	Rel	16	1		
		slow	other		
CHILDREN	A	14	6	.48	>.05
	Par	13	7		
	Pass	14	6		
	Rel	10	3		
		cold	bread		
EAT	A	8	6	13.70	<.005
	Par	6	6		
	Pass	2	8		
	Rel	0	15		
		bread	other		
COLD	A	19	1	7.78	>.05
	Par	20	0		
	Pass	18	2		
	Rel	15	5		
		cold	other		
BREAD	A	18	2	2.22	>.05
	Par	15	5		
	Pass	17	3		
	Rel	14	4		

cedure was repeated for the two other sentences. For all groups in Exp. II, no restriction was made on the possible responses during word association; any word from the sentence just learned was an acceptable response.

Results and Discussion

It was predicted that since the deep structures of all these sentences are essentially the same, no differences would be found in the patterns of associations in the different groups. The results of the word association tests are presented in Table 4. A word of explanation is in order concerning the entries in Table 4. The words from the sentence "SLOW CHILDREN EAT COLD BREAD" are used in the table to stand for the corresponding words in all three sentences. That is, SLOW stands for the three initial adjectives (SLOW, KNOWN, and HONEST), CHILDREN stands for the three subjects of the sentences, etc. The entry in any cell represents the number of *S*s out of the 20 in each group who gave the response in question for at least two of the three sentences. For example, the first row in Table 4 indicates that 19 *S*s in Group A gave the subject of the sentence as a response to at least two of the three SLOWS. Also, the very last line indicates that 14 *S*s of the 20 in Group rel gave COLD (i.e., the second adjective in the sentence) as a response to at least two of the three presentations of BREAD (i.e., the final noun in the sentence). This method of scoring results in each *S* contributing only one summarized score for each comparison.

The results were generally as predicted. For all groups, the most frequent response to each adjective was its corresponding noun; for each noun, the most frequent response was its corresponding adjective. Therefore, surface changes did not affect the adjective–noun pairs in any of the groups. (Although these results entail acceptance of the null hypothesis, it is felt that since they are presented only in conjunction with the significant differences found in Exp. I and not as a solitary finding, they are worth reporting.) However, a difference was found in the patterns of responses to EAT in the different groups. In the "pass" and "rel" sentences, more *S*s gave *bread* as a response to EAT, while in Groups A and par, about the same number of *S*s gave *bread* and *cold*. It may be that both deep and surface structures played a part in determining the response in this case. *Cold* would seem to be an acceptable response on the basis of the surface structure, since the whole phrase "cold bread" is the object of the verb in the surface structure in the "A" and "par" sentences.

GENERAL DISCUSSION

The two experiments reported lend support to the contention that the deep structures of sentences are relevant to sentence organization on a psychological level. In Exp. I, identical strings of words with different underlying structures were organized according to those structures, while in Exp. II, it was found that sentences with different surface structures but similar deep structures gave very similar patterns of associations. The present results can be thought of as further evidence to support the hypothesis that as soon as a speaker has dealt with a sentence, he is no longer responding to the surface structure.

However, Rosenberg (1966, 1968) has recently reported results that he felt suggested that factors besides underlying structure can influence the organization of sentences. He found that when *S*s attempted to recall sentences that had little or no interword association strength, recall seemed to be based on underlying structure. But if sentences were used that had strong associative connections among the words, the recall pattern was found to be significantly different than that based on underlying structure alone. No constituent breaks were found in the recall; the sentences seemed to be recalled as integrated units. Nevertheless, it should be noted that Rosenberg's interpretation of these data is open to question. The organization of the various sentences was only assessed through subsequent recall, not through any more direct method, such as the one reported here. It may be that the associative factor influences only recall and not the organization originally given to the sentence by *S*. An experiment has recently been completed in which the organization of some of Rosenberg's sentences was assessed through intrasentence word associations. Preliminary analysis of these data suggests that contrary to Rosenberg's theorizing, the high-associative sentences do not produce patterns of intrasentence word associations that deviate significantly from the patterns based on deep structure. Both the high- and low-associative sentences are organized in the same way. When S tries to recall a partially forgotten sentence, then associative factors may enter in terms of response availabilities and the like. A similar hypothesis has been proposed by Martin, Roberts, and Collins (1968). Associative factors would not seem to affect the organization given to sentences as they are learned and stored.

REFERENCES

AMMON, P. The perception of grammatical relations in sentences: A methodological exploration. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 869–875.

BLUMENTHAL, A. Prompted recall of sentences. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 203–206.

CHOMSKY, N. *Aspects of the theory of syntax.* Cambridge: M. I. T. Press, 1965.

COFER, C., & BRUCE, D. Form-class as the basis for clustering in the recall of nonassociated words. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 386–389.

GARRETT, M., BEVER, T., & FODOR, J. The active use of grammar in speech perception. *Perception and Psychophysics,* 1966, **1**, 30–32.

JOHNSON, N. The influence of associations between the elements of structured verbal responses. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 369–374.

MARTIN, E., ROBERTS, K., & COLLINS, A. Short-term memory for sentences. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 560–566.

ROHRMAN, N. The role of syntactical structure in the recall of English nominalizations. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 904–912.

ROSENBERG, S. Recall of sentences as a function of syntactic and associative habit. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 392–396.

ROSENBERG, S. Association and phrase structure in sentence recall. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 1077–1081.

SACHS, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics,* 1967, **2**, 437–442.

SUCI, G., AMMON, P., & GAMLIN, R. The validity of the probe latency technique for assessing structure in language. *Language and Speech,* 1967, **10**, 69–80.

(Received March 10, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 339-342

DIFFERENTIAL RATES OF GSR HABITUATION TO PLEASANT AND UNPLEASANT SAPID STIMULI[1]

G. LAWRENCE FISHER[2] AND BARBARA E. FISHER[3]

University of Maryland

A study was done to compare diminution of the GSR to repetitive stimulation by different concentrations of hedonically positive (sucrose), negative (quinine monohydrochloride), and neutral (water) solutions. Each of nine groups of *S*s were stimulated by a single concentration of one of the solutions. Habituation of the GSR to the quinine required significantly more presentations than to sucrose or water. Since the orienting reaction sensitizes the individual to potential trauma and since most noxious compounds are potentially toxic, it follows that the GSR component of the orienting reflex (OR) should be slower to habituate to quinine than the other solutions. This hypothesis was supported by obtained data. Quinine concentration was a significant factor, while sucrose concentration was not. The findings were extended to provide a preliminary evaluation of a conjectured gustatory affect process that is responsible for the repugnance shown when bitter compounds are tasted.

Sensory stimuli may perform two functions: informational and motivational. Although most investigators of sensory processes stress the informational one, the two functions are related. This is particularly apparent in gustation. In the present study the investigators attended to the motivational process, i.e., to the role of perceptual taste cues in the evocation of autonomic, defensive responses.

Most mammals reject substances reported as "bitter" and accept those reported as "sweet" by humans. Both bitter and sweet substances are found in nearly every chemical class (Moncrieff, 1967). Furthermore, there is an abundance of evidence showing that bioelectrical activity of the classic gustatory structures varies with the intensity of stimulation, regardless of the acceptance–rejection characteristics of the chemical (Pfaffmann, 1965; Pfaffmann, Fisher, & Frank, 1967).

The nearly universal repugnance for bitter chemicals suggests the existence of an innate, plastic process whose primary function is to reject aversive substances. The present study was done to evaluate this suggestion. The orienting reflex (OR) is believed to reflect a mobilization of the organism's bodily resources for action on impinging stimulation (Sokolov, 1963). The GSR component of the OR was selected for use in this study because, despite the lack of consensus regarding its origins, it is generally agreed that it is an indication of sympathetic arousal (Woodworth & Schlosberg, 1954). Furthermore, Sokolov (1963) and others (Montagu & Coles, 1966; Uno & Grings, 1965; Zimny & Schwabe, 1966) have confirmed that the GSR diminishes with repetitive sensory stimulation. If a component of the OR habituates differentially to hedonically positive and negative sapid solutions, one may systematically study the proposed process that may be responsible for taste-dependent food aversions or rejection.

METHOD

Subjects.—Ninety-three introductory psychology students at the University of Maryland served as *S*s to meet a course requirement. The data of three *S*s were discarded because of equipment failure. Five males and five females were assigned to each of nine groups. Four of these groups were stimulated with sucrose solutions, four with quinine, and one with water.

---

[1] A portion of the data in this study was obtained for an honors thesis by the junior author under the supervision of the senior author. The work was supported by National Science Foundation Grant GB-5843.

[2] Requests for reprints should be sent to G. Lawrence Fisher, Department of Psychology, University of Maryland, College Park, Maryland 20742.

[3] Now at Pennsylvania State University, Milton S. Hershey Medical Center.

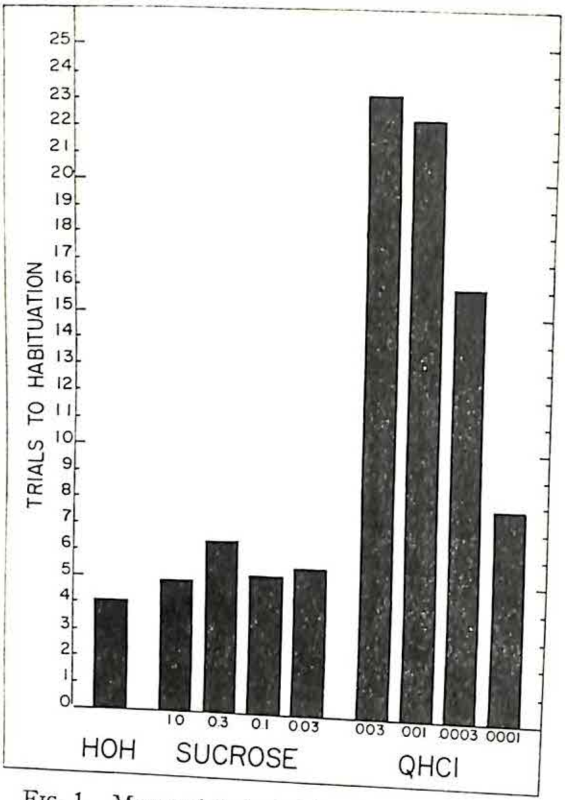


FIG. 1. Mean trials to habituation for 90 *S*s, 10 in each Chemical × Concentration. (Numbers on abscissa are M/l.)

Apparatus.—The stimulus solutions were delivered via a solenoid valve and nozzle controlled by *E*. A thistle tube above the valve held the solution until released. The thistle tube was filled using a battery of refillable syringes; each syringe supplied either deionized water or a single concentration of sucrose or quinine monohydrochloride (QHCl).

The stimulus solutions were: quinine monohydrochloride (USP)—.003M, .001M, .0003M, .0001M; sucrose (reagent grade)—1.0M, .3M, .1M, .03M; and deionized water.

The GSR and BSR were recorded using Ag-AgCl electrodes. A circular electrode, approximately 2 cm.^{2} in area, was applied to the middle phalanx of the middle finger of the left hand with masking tape. The reference electrode (50 × 75 mm.) was applied to the glabrous surface of the ipsilateral, upper left arm. K-Y surgical jelly (Johnson and Johnson) served as an electrolyte. The exosomatic responses were amplified using a Model 201 Biophysical Research Instruments GSR/BSR amplifier with a 20-μa. constant current supply. Responses were recorded on a dual-channel Texas Instrument Model 302 galvanometric recorder.

The *S* was seated in a comfortable dental chair within an electrically shielded room. After the electrodes were applied, the chair was elevated until *S*'s tongue could be extended comfortably beneath a short oxygen catheter connected to the solenoid valve orifice. A heat-molded, polyethylene funnel was brought against the chin immediately below the lower lip. The solution flowed over the anterior dorsal tongue into the funnel. The *S* maintained his tongue in the extended position for the entire session. Interstimulus intervals were variable, the shortest being 30 sec., the longest, 75 sec.

The room was darkened and "Noisefoes" were placed over *S*'s ears to reduce extraneous, auditory stimulation. The solenoid valve was packed in plastic foam within a grounded aluminum box. The only nongustatory cues were tactile and thermal when the solution flowed over the tongue. All solutions were maintained at 22–25° C.

Each *S* was stimulated repetitively with 10 ml. of one concentration of a solution until either the habituation criterion was reached or 26 presentations were completed. The criterion for a stimulus-related response required that it occur within 1–4 sec. after the solution was presented and that it was clearly distinguishable from background, nonspecific activity. The smallest skin resistance change accepted as a specific response was 145 ohms. The session length was limited to 1 hr., which permitted stabilization of the BSR (10–15 min.) and completion of recording before *S* became uncomfortable.

RESULTS

The number of trials to habituation was counted for each session. The requirement for habituation was that no response occurred for two sequential trials. Figure 1 summarizes the major finding. The GSR component of the OR is more resistant to habituation to QHCl than to sucrose or water. Furthermore, the habituation is related to the concentrations of the QHCl solutions, while unrelated to the concentrations of sucrose.

The sapid chemical effect was significant, $F(1, 27) = 86.6$, $p < .001$. The concentration effect was also significant, $F(3, 27) = 8.0$, $p < .005$. The Solution × Concentration interaction was significant, $F(3, 27) = 7.7$, $p < .001$. Variation among *S*s was not significant. Figure 1 shows clearly that the significant concentration effects and Concentration × Solution interactions were contributed by the QHCl.

All *S*s habituated to water by the fifth trial. Of 40 *S*s, 39 habituated to sucrose by the tenth stimulation; a single *S* continued to respond for 18 trials to .3M sucrose. Of

40 *S*s, 18 did not habituate to QHCl by the twenty-fifth trial. All responses were transformed to changes in conductance, $\sqrt{\Delta C}$ based on the formula proposed by Edelberg (1967):

$$\Delta C = \frac{GSR}{(BSR)^2}$$

The mean conductance was determined on a trial-by-trial basis for each chemical. An analysis of variance indicated a lack of significance among concentrations of sucrose and QHCl on Trials 1 and 2. The total number of responses was the same for these trials, i.e., five *S*s habituated by the third trial.

Discussion

Using Sokolov's (1963) theory as a framework, it was hypothesized that GSRs to QHCl would habituate more slowly than sucrose for the following reasons: (*a*) In many instances, bitter substances are toxic, and the organism reacts defensively to their presentation; after having determined the chemical's stimulus properties, he will generally avoid intake. This defensive process requires a high level of arousal, thereby perpetuating the OR. (*b*) Sweet substances are generally beneficial to the animal; once he has determined its stimulus characteristics, the need for a high level of arousal is removed, i.e., whether or not intake occurs. (*c*) Sugar, in any of its forms, is a more common element in the typical diet than quinine. Rapid habituation to sugar may be partially induced by familiarity.

The present experimental results were consistent with these expectations. While many *S*s were still responding quite strongly to quinine on the twenty-fifth trial, all habituated to sugar and water prior to the tenth trial. These results are relatively conservative since (*a*) the mean interstimulus interval for quinine was shorter (36.3 sec.) than that for sucrose (43.3 sec.), thus facilitating a relatively more rapid habituation to the former (Allen, Hill, & Wickens, 1963; Thompson & Spencer, 1966), and (*b*) only the anterior, or more sucrose-sensitive, portion of the tongue was stimulated (Amerine, Pangborn, & Roessler, 1965).

Of 10 *S*s, 8 were still responding on the twenty-fifth trial at the two higher concentrations of QHCl. The lower concentrations of QHCl produced more variable data. The lowest concentration was reported as not being particularly unpleasant by the present *S*s. In some cases, the lowest concentrations of the QHCl and sucrose were reported informally as barely recognizable.

The water-elicited GSR required an average of 4.1 presentations to habituate. Consider that two factors are operating: habituation to the situation (nonspecific) and habituation to the stimulus itself (stimulus-specific habituation). The *S*s usually respond to the first presentation of any unexpected stimulus. In the present study, *S*s were experimentally naive and generally apprehensive about serving as *S*s in an experiment involving surface electrodes and a dental chair. Once the novelty and trepidation dissipated, the nonspecific habituation progressed. The sucrose required more stimulus presentations to habituate than water, but this effect was not significant, even at the highest concentration.

One problem raised by these data is that of identifying the causes of the different rates of habituation. One may argue that the GSR habituation to various solutions was due to the adaptation of subjective intensity of the solution rather than to an affective process (Krakauer & Dallenbach, 1937; Meiselman, 1968). This is unlikely; Krakauer and Dallenbach found that complete adaptation of QHCl and sucrose requires 75–325 sec. of continuous stimulation. Slight tongue movements prolong adaptation. In the present study, each sapid solution flowed over the tongue for about 6 sec/trial (150 sec. for 25 trials). This is well below the adaptation time required for the strongest solution in the experiment cited. In addition, the interstimulus intervals were longer than 30 sec., during which time small, lingual movements occurred. During debriefing, most *S*s reported that stimulus intensity remained fairly constant.

On the other hand, the data may reflect qualitative differences in gustatory sensation; i.e., they may reflect the hedonic characteristics of the chemicals apart from the taste name, e.g., *disgusting* vs. *bitter*. As a supporting example, sodium chloride is often reported unpleasant at hypotonic concentrations and unpleasant at hypertonic concentrations (Pangborn, 1967). A study similar to the current one, using weak and strong concentrations of the same chemical, would emphasize the affective dimension of the stimulus complex. If future results indicate that autonomic response habituation reflects differences in the affectivity of stimuli, this additional parameter may have to be specified in those habituation studies in which the stimuli have significant affective components.

Stein (1966) has postulated that habituation is the result of an interaction between excitatory and inhibitory processes. Following stimulus onset, the excitatory process is activated and then the inhibitory one. The latter may be conditioned to override the excitatory process upon stimulus onset, thereby diminishing the OR, i.e., habituating. This is an interesting view. Taste stimuli are always accompanied by nongustatory cues. There is ample opportunity for these cues to assume the hedonic attributes of the taste itself.

During informal taste tests in our laboratory, distilled water was reported as "bitter," precluding its use as a neutral or tasteless solution. On the other hand, deionized water was reported "tasteless" or good because it didn't have any flavor. The ionic content of the solvent and rinse was less than 1 ppm in terms of NaCl conductivity.

The concentrations of QHCl and sucrose were selected on the basis of subjectively equal intensities (Fisher;[4] Kocher & Fisher, 1969). The .003M QHCl was reported as less intense than the 1.0M sucrose; the .001M QHCl was reported as less intense than the .3M sucrose. The lowest concentrations of chemicals used in the present study were generally reported as "slightly bitter" or "slightly sweet" in the Kocher and Fisher study. It may be assumed that the differences in GSR resistance to habituation shown in the present study are primarily attributable to factors other than intensity alone. At the higher concentrations, the QHCl was reported markedly "unpleasant," while the sucrose was "pleasant" in the studies. Both solutions were colorless, odorless, and of approximately equal viscosities and temperature (22–25° C.) at the time of presentation. The direction of the hedonic tone of these chemicals is the most prominent distinguishing characteristic. These data do not contradict the possibility that there is a secondary gustatory process whose primary function is to react differentially to hedonic attributes of the taste stimulus complex.

---

[4] Unpublished study entitled "Magnitude estimations of intensity and preference of quinine and sugar solutions," 1969.

REFERENCES

Allen, C. K., Hill, F. A., & Wickens, P. O. The orienting reflex as a function of the interstimulus interval of compound stimuli. *Journal of Experimental Psychology*, 1963, **65**, 309–316.

Amerine, M. A., Pangborn, R. M., & Roessler, E. B. *Principles of sensory evaluation of food.* New York: Academic Press, 1965.

Edelberg, R. Electrical properties of the skin. In Clinton C. Brown (Ed.), *Methods in psychophysiology.* Baltimore: Williams & Wilkins, 1967.

Kocher, E., & Fisher, G. L. Subjective intensity and taste preference. *Perceptual and Motor Skills*, 1969, **28**, 735–740.

Krakauer, D., & Dallenbach, K. M. Gustatory adaptation to sweet, sour, and bitter. *American Journal of Psychology*, 1937, **49**, 469–475.

Meiselman, H. L. Magnitude estimations of the course of gustatory adaptation. *Perception and Psychophysics*, 1968, **4**, 193–196.

Moncrieff, R. W. *The chemical senses.* London: Leonard Hill, 1967.

Montagu, J. E., & Coles, E. M. Mechanism and measurement of the galvanic skin response. *Psychological Bulletin*, 1966, **65**, 261–279.

Pangborn, R. M. Some aspects of chemoreception in human nutrition. In M. R. Kare & O. Maller (Ed.), *The chemical senses and nutrition.* Baltimore: Johns Hopkins University Press, 1967.

Pfaffmann, C. De Gustibus. *American Psychologist*, 1965, **20**, 21–31.

Pfaffman, D., Fisher, G. L., & Frank, M. K. The sensory and behavioral factors in taste preference. In T. Hayashi (Ed.), *Olfaction and taste: II.* New York: Pergamon Press, 1967.

Sokolov, E. N. *Perception and the conditioned reflex.* Oxford: Pergamon Press, 1963.

Stein, L. Habituation and stimulus novelty: A model based on classical conditioning. *Psychological Review*, 1966, **73**, 352–356.

Thompson, R. F., & Spencer, W. A. Habituation: A model phenomenon for the study of neural substrates of behavior. *Psychological Review*, 1966, **73**, 16–43.

Uno, T., & Grings, W. W. Autonomic components of orienting behavior. *Psychophysiology*, 1965, **1**, 311–321.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, Rinehart & Winston, 1954.

Zimny, G. H., & Schwabe, L. W. Stimulus change and habituation of the orienting response. *Psychophysiology*, 1966, **2**, 103–115.

(Received March 12, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 343–346

SEMANTIC SIMILARITY AND THE COMPARISON OF WORD MEANINGS[1]

BENSON SCHAEFFER[2] AND RICHARD WALLACE

University of Oregon

Two experiments attempted to demonstrate the effects of semantic similarity on the judgment of word meanings. Experiment I showed that semantic similarity facilitated the judgment of meaning equivalence, e.g., the judgment that two words belong to the same category. Experiment II showed that semantic similarity hindered the judgment of meaning difference, e.g., the differentiation between the meanings of two words. Explanations of the effects in terms of associative biases and hierarchical processing are proposed.

Semantic similarity has been shown to affect performance in many situations. Razran (1949) found that a conditioned salivary response generalized on the basis of semantic similarity, and Corman and Wickens (1968) showed that increasing the semantic similarity between interpolated and recalled material decreased the retention of the latter in short-term memory. Recently, Posner and Keele (1967), Neisser (1967), and Landauer and Freedman (1968) have used reaction time tasks to investigate stages in perceptual processing, parallel and serial processes, and long-term memory search, respectively. The present experiments employ reaction time tasks to highlight the effects of semantic similarity on the judgment of word meanings.

Word meanings can be compared along dimensions of equivalence or difference; e.g., the words *wren* and *fox* are equivalent in that both denote an animal and different in that only one denotes a mammal. The present studies attempt to show that semantically similar words can be more easily judged equivalent in meaning than can semantically dissimilar words, but less easily judged different. The prediction is based on the assumption that associations are activated when a given word is understood. Two semantically similar words can be more readily judged equivalent than can dissimilar ones because they activate many of the same associations and these provide cues of similarity. Two similar words can be less readily judged different because the cues of similarity provided by their common associations interfere with the difference judgment. Experiments I and II attempt to demonstrate these effects.

EXPERIMENT I

A same–different reaction time task was used to find out if semantic similarity facilitates equivalence judgments. The *S*s were shown a pair of words and asked to judge whether the two words had the same or different meanings. They were instructed to press the "same" button if both words denoted living things or if both denoted nonliving things, and to press the "different" button if one denoted a living thing and the other a nonliving thing. The "living" words were mammals and flowers and the "nonliving" words were metals and fabrics. It was predicted that *S*s would press the "same" button more quickly to two mammals or two flowers than they would to a mammal and a flower; and more quickly to two metals or two fabrics than to a metal and a fabric.

Method

Subjects.—Five male and five female students at the University of Oregon served individually as *S*s in two sessions on consecutive days and were paid $1.50 per hour.

Materials.—The presentation materials were 320 pairs of nouns. The pairs were derived from a set of 40 words ordered as follows: 20 were living words, divided further into 10 mammals and 10 flowers; and 20 were nonliving words, divided

---

[1] This research was supported by the Advanced Research Projects Agency, United States Department of Defense, and was monitored by the United States Air Force Office of Scientific Research under Contract F44620-67-C-0099.

[2] Requests for reprints should be sent to Benson Schaeffer, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

TABLE 1

MEAN REACTION TIME (IN SEC.) FOR SAME AND DIFFERENT JUDGMENTS ON DAY 1 AND DAY 2

Judgment	Day 1	Day 2
Mammal-Flower-Same	1.523	1.324
Flower-Flower-Same	1.431	1.187
Mammal-Mammal-Same	1.315	1.152
Metal-Fabric-Same	1.887	1.522
Fabric-Fabric-Same	1.774	1.407
Metal-Metal-Same	1.447	1.192
Living-Nonliving-Different	1.655	1.422

further into 10 metals and 10 fabrics. Living words had either two or three syllables (five of each in each subgroup) and nonliving words had one or two syllables (with the same allotments). The words were: *lion, squirrel, camel, raccoon, zebra, antelope, gorilla, kangaroo, elephant,* and *opossum; tulip, pansy, iris, lily, daisy, hyacinth, petunia, carnation, marigold,* and *gardenia; gold, lead, tin, zinc, steel, iron, copper, silver, nickel,* and *tungsten;* and *wool, silk, felt, tweed, yarn, flannel, denim, rayon, nylon,* and *velvet.*

The word pairs were of three types: in 80 pairs both words were members of the same subgroup (mammal, flower, metal, or fabric); in 80 pairs each word was a member of a different subgroup, but both belonged to the same superordinate group (living or nonliving); and in 160 pairs each word belonged to a different superordinate group. The values of the following variables were balanced for each type: syllable number, subgroup membership, and appearance of a word on the left or right.

Each word pair was typed in capital letters (⅛ in. high) on one line of a 5 × 8 in. card, with the words and the space between them centered 1½ in. from the bottom edge.

The cards were shown in a tachistoscope. When *S* was leaning into the eyepiece, the card was approximately 18¼ in. from his eyes, and the word pairs subtended a visual angle of about 4°. The *S*s sat in a chair in front of the tachistoscope; attached to the chair was a horizontal panel with two response keys, one labeled "same," one "different." When the tachistoscope field was exposed, depression of one of the keys turned it off and stopped a Hunter timer which registered response time in milliseconds. During the sessions, *S* wore headphones which transmitted white noise at a constant low amplitude.

Procedure.—The set of 320 cards was divided into two decks of 160. Each deck was balanced with respect to the variables mentioned previously. Each deck was ordered randomly, with the constraints that in one half of the pairs the words belonged to different superordinate categories and that there were no more than four same- or different-superordinate pairs in a row. The two resulting orders were used for all *S*s.

The *S*s were run in two sessions held on consecutive days. In each session, one of the decks was used, yielding 160 trials. Five *S*s were given one deck first, and five the other.

At the beginning of his first session, *S* was told that his task was to decide, on each trial, whether the two words presented were both living, or nonliving, or that one word belonged to each of these categories. If both words belonged to the same category, he was to press the key labeled "same"; if they did not, he was to press the key labeled "different." The *S* was told that the living words would refer to mammals or flowers and the nonliving words to metals or fabrics. Finally, he was told that this was a reaction time experiment, and therefore he was to respond as quickly as possible while still keeping his errors to a minimum.

In each session the initial two trials were "different" pairs; the first of these was not counted in the analysis. The series of trials in a session was divided into groups of 15 and a rest period of about 45 sec. was given between groups. Within each group the intertrial interval was about 5 sec. Immediately before each trial, *E* said "ready" and then presented a card. A session lasted about 1¼ hr.

Results

The mean error rate was 7.8%, low enough to consider correct reaction times a good measure of judgment difficulty. Table 1 presents the mean reaction time in seconds for correct same and different judgments on Day 1 and Day 2.

As can be seen, mammal-flower-same judgments were more difficult than either flower-flower- or mammal-mammal-same judgments, and metal-fabric-same judgments were more difficult than either fabric-fabric- or metal-metal-same judgments. A repeated measures analysis of variance showed that the average difference, .176 sec., was significant at the .001 level, $F\ (1, 54) = 70.00$. The smallest difference, between metal-fabric- and fabric-fabric-same judgments, was significant at the .01 level, $F\ (1, 54) = 7.31$. Clearly, semantic similarity facilitates judgments of meaning equivalence.

The session or practice effect was significant at the .01 level, $F\ (1, 9) = 14.79$, and the Practice × Judgment interaction was significant at the .05 level, $F\ (6, 54) = 3.32$. The latter was a result of the smaller difference between mammal-mammal- and flower-flower-same judgments on Day 2 than on

Day 1. There were also consistent differences in difficulty as a function of semantic category.

EXPERIMENT II

A modified same–different reaction time task was used to find out if semantic similarity hindered the judgment of meaning difference. The *S*s were presented with two category labels, such as bird and mammal, followed by a single word, such as *wren*, that belonged to one of the two labeled categories. The *S*'s task was to press the button corresponding to the side, left or right, on which the category label which described the single word had appeared. (The correct position varied randomly.) The question of interest was: would the judgment of a word, such as *wren*, be quicker when two semantically similar labels, such as bird and mammal, appeared or when two semantically dissimilar labels, such as bird and fruit, appeared. It was predicted that the common associates, such as *animal*, of the similar labels, bird and mammal, would hinder the differentiation between them. The labels bird, mammal, fruit, and grain were used.

Method

Subjects.—Five male and five female students at the University of Oregon served individually as *S*s in two sessions on successive days and were paid $1.50 per hour.

Materials.—The presentation materials were constructed from a list of four category labels (bird, mammal, fruit, and grain) and 20 item words, 5 of which belonged to each category (*wren*, *dove*, *robin*, *hawk*, and *duck*; *fox*, *deer*, *tiger*, *mouse*, and *mule*; *orange*, *plum*, *grape*, *pear*, and *apple*; and *rice*, *oats*, *barley*, *wheat*, and *rye*).

Pairs of category words, including all 12 permutations, and single item words were typed in capital letters (⅛ in. high) on white 5 × 8 in. cards, centered 1½ in. from the bottom edge.

The cards were presented in a two-channel tachistoscope. The tachistoscope, *S*s' seating and response panel, and the way in which response time was recorded were the same as in Exp. I.

Procedure.—The *S*s participated in two 1½-hr. sessions on consecutive days. In each trial of a session, *S* was first presented with a pair of category words for 700 msec., followed immediately by an item word that belonged to one of the two categories. This word remained in view until *S* made his response. The *S*'s task was to decide to which category the item word belonged and to press the key corresponding to the position of the correct category word; e.g., if the correct category word was the left word of the category pair, *S* had to press the left key. His response time was measured from the onset of the item word to the depression of a response key.

All combinations of category-word pairs and relevant item words appeared in each 120-trial session. Trials were run in groups of 10 with a rest of about 45 sec. between groups. Within a group, the intertrial interval was about 10 sec. Twelve practice trials were given prior to Session 1, and two prior to Session 2.

TABLE 2

MEAN REACTION TIME (IN SEC.) FOR JUDGMENTS CONCERNING SEMANTICALLY SIMILAR AND DISSIMILAR LABELS

Similar labels		Dissimilar labels			
Bird-Mammal	Fruit-Grain	Bird-Fruit	Bird-Grain	Mammal-Fruit	Mammal-Grain
.784	.774	.742	.748	.723	.731

Results

The mean error rate was 7.5%, low enough to consider correct reaction times a good measure of judgment difficulty. Table 2 presents the mean reaction time in seconds for correct judgments concerning similar and dissimilar labels.

As can be seen, semantically similar labels were harder to differentiate than were semantically dissimilar ones. A repeated measures analysis of variance showed that the average difference of .043 sec. was significant at the .001 level, $F\ (1,\ 45) = 51.49$. The smallest difference, between grain-fruit and grain-bird judgments, was significant at the .05 level, $F\ (1,\ 45) = 7.15$. Clearly, similarity hindered judgments of meaning difference. The session or practice effect was also significant at the .01 level, $F\ (1,\ 9) = 14.32$.

DISCUSSION

The experiments showed that semantic similarity facilitates the judgment of meaning equivalence, but hinders the judgment of meaning difference. A consideration of the individual experiments will help refine the associative interpretation of the results put forth in the introduction. The data of Exp. I can be explained in two ways. An associative bias interpretation would hold that the associations common to similar words evoked an implicit "same" response that facilitated the "same" button press; a hierarchical processing interpretation would hold that two similar words can be judged equivalent in fewer steps than can two dissimilar ones; e.g., *tulip* and *pansy* would be judged "flower-flower" and "same," while *iris* and *camel* would be judged "flower-mammal," "living-living," and "same." The data do not distinguish between interpretations. Posner and Keele (1967) found an analogous result at the perceptual level; e.g., two physically identical letters, A and A, were more quickly judged name-identical than were two dissimilar letters, A and a.

The results of Exp. II can be explained in many ways. Recall that it was harder to call a wren a bird when the label *bird* appeared with the label *mammal* than when it appeared with the label *fruit*. Two forms of associative-bias interpretation suggest themselves. One holds that the two labels, *bird* and *mammal,* arouse a common associate, *animal,* which evokes an implicit "same" response that interferes with the judgment. The other holds that the label and the word, *mammal* and *wren,* evoke the common associate, *animal,* that interferes with the judgment. Both interpretations are likely true in part. What is clear, however, is that a hierarchical processing interpretation of the type proposed for the results of Exp. I will not do. The basic reason is that the hierarchical processing posited is an efficient procedure that minimizes the number of steps in the judgment process. The judgment process taking place in Exp. II is not an efficient procedure because it involves the reading of both labels. Since one of the labels always describes the word, *S*s would be efficient if they read only one label, say the one on the left, and pressed the left button if this label described the word and the right button if it did not. Instead, as the results show, they read both labels. Henmon (1906) found a result analogous to that of Exp. II at the perceptual level; e.g., two similar colors, red and yellow, were less quickly judged different than were two dissimilar colors, red and green.

REFERENCES

Corman, C. D., Wickens, D. D. Retroactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1968, 7, 16–19.

Henmon, V. A. C. The time of perception as a measure of difference in sensation. *Archives of Philosophy, Psychology and Scientific Methods,* 1906, No. 8.

Landauer, R. K., & Freedman, J. L. Information retrieval from long-term memory: Category size and recognition time. *Journal of Verbal Learning and Verbal Behavior,* 1968, 7, 291–295.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Posner, M. I., & Keele, S. W. Decay of visual information from a single letter. *Science,* 1967, 158, 137–139.

Razran, G. H. S. Semantic and phonetographic generalizations of salivary conditioning to verbal stimuli. *Journal of Experimental Psychology,* 1949, 39, 642–652.

(Received March 13, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 347–352

DOUBLE STIMULATION WITH VARYING RESPONSE INFORMATION[1]

BARRY H. KANTOWITZ[2]

University of Wisconsin

The S_1-R_1-S_2 double-stimulation paradigm was applied to a two-choice Fitts' law task. Fitts and Peterson concluded that in a single-stimulation S_1-R_1 task, reaction time (RT) and movement time (MT) reflect independent processes. Parallel Fitts' law functions were obtained for both single- and double-stimulation conditions. However, in the double-stimulation conditions, MT was slightly elevated relative to the single-stimulation condition. RT in the double-stimulation conditions was considerably greater than single-stimulation RT. It was concluded that RT and MT reflect additive rather than independent stages of information processing, and an incremental two-stage model was proposed.

Views of a man as a serial information processor have been supported by the relative independence of selection and execution stages of motor responses in a choice reaction time (RT) task. Fitts and Peterson (1964) investigated effects of varying response information on both RT and movement time (MT) in a discrete two-choice motor task. The *S*s were required to move as quickly as possible to one of two alternative targets; however, both the target width (W) and the distance to the target (A = amplitude of movement) were varied. The amount of information generated by the response was given by an index of difficulty (ID) proposed by Fitts (1954) and equals $\log_2 (2A/W)$ bits. Fitts, as well as Fitts and Peterson, found MT to be a linear function of ID; this relationship has been often replicated and is generally termed Fitts' law. However, RT was not found to be affected by ID. Thus Fitts and Peterson concluded that RT indexes response selection, while MT is a measure of response execution, and that these processes are relatively independent.

Also compatible with a distinction between response selection and response execution are results of two related double-stimulation studies (Herman & Israel, 1967; Herman & Kantowitz, 1969). In the general double-stimulation paradigm, two successive stimuli are presented in close temporal proximity, e.g., less than approximately 500 msec. apart. Response demands vary with the particular task chosen (for a recent discussion, see Herman & Kantowitz, in press). The paradigm of present interest can be schematized as S_1-R_1-S_2. Note that no overt response is required for the second stimulus. This paradigm was used by Helson and Steger (1962), who found RT to S_1 to be increased when the S_1-S_2 interstimulus interval (ISI) ranged from 10 to 170 msec. Unsuccessful replications by Lappin and Eriksen (1964) and Koplin, Fox, and Dozier (1966) were attributed by the authors to methodological flaws in the Helson and Steger study. This unsatisfactory state of affairs was clarified by Herman (1969), who noted differences between the stimulus display used by Helson and Steger and that used in the replication attempts. Both Lappin and Eriksen and Koplin et al. used a stimulus display comprised of two lights alternating as S_1 and S_2 on successive trials; thus each light served a dual role, functioning (on different trials) as either S_1 or S_2. The display used by Helson and Steger consisted of three lights, with the two end lights serving as S_1 and the third central light always as S_2; thus

---

[1] This research was conducted while the author was supported by United States Public Health Service Predoctoral Fellowship 7-F1-MH-29. The author is indebted to E. E. Smith for his comments on the manuscript and to Bob Calfee, David Grant, and Leonard Ross for assistance in obtaining equipment. Additional criticism by the editorial consultant and by Louis Herman is also acknowledged.

[2] Requests for reprints should be sent to Barry H. Kantowitz, who is now at Purdue University, Department of Psychology, Lafayette, Indiana 47907.

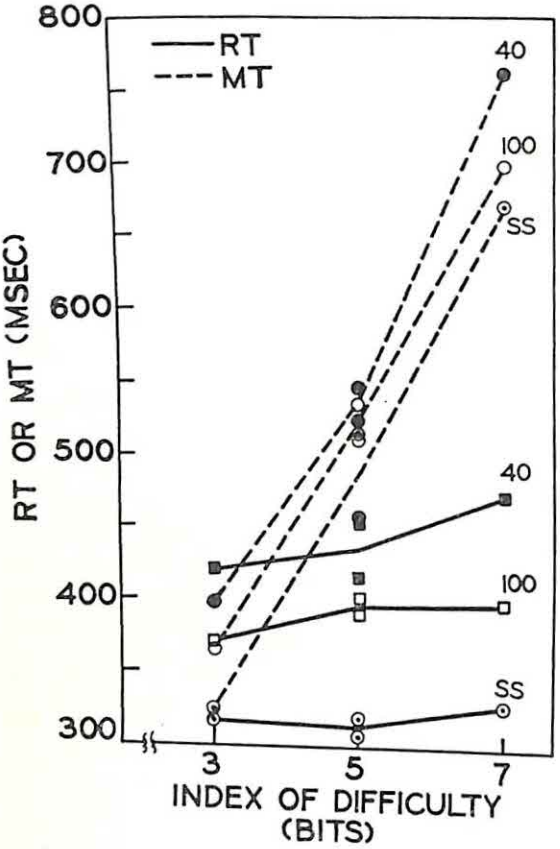


FIG. 1. Mean latencies as a function of ID. (SS = single-stimulation control condition; 40 = 40-msec.-ISI, double-stimulation condition; 100 = 100-msec.-ISI, double-stimulation condition. All points are based on errorless trials.)

The main effect of ISI was significant, F (2, 22) = 13.13, $p < .001$, but the ID × ISI interaction was not, F (6, 66) = 1.42, $p >$ .05. Therefore, the three Fitts' law functions can be considered parallel. There was a slight but significant elevation in MT for the double-stimulation conditions as compared to the single-stimulation control condition. Two-tailed multiple t tests were used to compare 40-msec. ISI with single-stimulation control, t (22) = 5.13, $p < .001$, 100-msec. ISI with single-stimulation control, t (22) = 2.88, $p < .01$, and 40-msec. ISI with 100-msec. ISI, t (22) = 2.26, $p < .05$, since no a priori theoretical justification existed for predicting a directional effect.

Since the 5-bit ID was achieved with two targets of different widths, Fig. 1 shows six points at that ID level as compared to three points at other ID levels. Fitts' law predicts no differences in MT between the two 5-bit target sets. An additional analysis of variance, two levels of ID and three levels of ISI, failed to reveal a significant difference between MTs for the two targets, F (1, 11) = 3.94, $.01 > p > .05$.

Reaction time.—Figure 1 also shows RT as a function of ID, with ISI as the parameter. The approximately horizontal functions indicate that RT was not influenced by ID, F (3, 33) = 1.35, $p > .05$. The main effect of ISI was significant, F (2, 22) = 11.94, $p < .001$. On the basis of both prior empirical findings and response conflict theory (Herman & Kantowitz, in press), directional multiple t tests were used to compare the single-stimulation control condition with the 40- and 100-msec. ISI double-stimulation conditions. The one-tailed comparisons between control and 100-msec. conditions, t (22) = 2.86, and between control and 40-msec. conditions, t (22) = 4.84, were significant at better than the .005 and .0005 levels, respectively. An additional one-tailed test between the two double-stimulation conditions evidenced reliably greater elevation at the shorter ISI, t (22) = 1.98, $p < .05$. The ID × ISI interaction was not significant, F (6, 66) = .83, $p > .05$.

The frequency of moving-to-the-wrong-target errors was greatest for the 40-msec. ISI (3.60 errors/block), followed by the 100-msec. ISI (.67 errors/block) and ∞ ISI (0 errors/block). Mean RT of these error trials was greater than RT for correct trials for both 100-msec. (mean error RT = 444 msec.) and 40-msec. (mean error RT = 582 msec.) ISIs.

DISCUSSION

Results are fully in accord with Predictions *a* and *b*, generated by an additive model. Although Prediction *c* of the independence model is supported, Prediction *d* is not, thus invalidating the independence assumption. Figure 2 shows two additive models, both of which are compatible with the data presented. In Model A the selection phase is influenced only by ISI, but the execution phase is a function of both ISI and ID. Sternberg (1968) has also used noninteracting independent variables to define a class of additive models similar to

Model A. However, Sternberg's analyses are based on total response time, which in the present study would be the sum RT + MT, so that his models are not directly comparable to Model A. Nevertheless, both types of models are similar in stating that human information processing involves a series of stages and that one stage must be completed before the next can begin. Model A implies an all-or-none response selection, with the output (but not the duration) of the selection stage being uninfluenced by ISI. This is followed by a two-stage execution phase. In contrast, Model B presents an execution phase consisting of but one stage. The ISI variable influences MT by inserting noise and degrading the output of the selection stage. This type of analysis diverges from that of Sternberg (1968, p. 12), whose definition of stage precludes the output of a stage being influenced by the same independent variable responsible for the duration of that stage. Model B implies that the probability of the emission of a particular response selection increases without discontinuity during the selection phase. To determine which model[4] is to be preferred, some inferences need to be drawn from RT data.

The present study found RT to be greater for the shorter (40-msec.) ISI. This is in agreement with results of Herman and McCauley (1969), who found RT to decrease over ISI values of 30, 60, 90, and 120 msec., with the greatest decrease occurring between 30 and 60 msec. Such a finding is entirely consonant with response conflict theory, which predicts greater conflict, and hence greater delay, as the competing response tendencies approach each other in time. However, this finding was only barely significant in the present study, and other studies (Herman, 1969; Herman & Israel, 1967) have found RT to remain constant, although delayed relative to a single-stimulation condition, over ISI values of 50, 100, and 150 msec. Thus it is reasonable to consider another indicant of response conflict. In the present study, error frequency was considerably greater at the shorter ISI, whereas RT was but slightly greater. Furthermore, RT was greater on error trials than on correct trials, although this evidence for increased conflict is not reflected

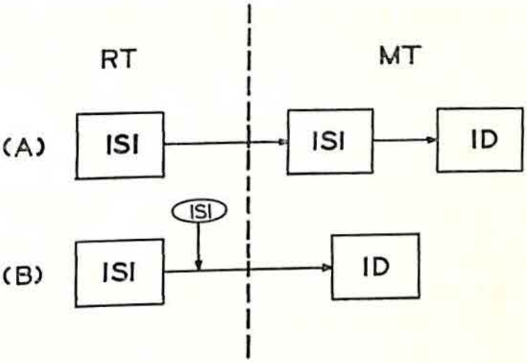


FIG. 2. Two possible information-processing models. (The dashed vertical line separates response selection on the left from response execution on the right.)

between RTs for the two ISI values since error trials, i.e., presumably those trials arising from maximum response competition, were excluded from analysis. Herman and McCauley, while reporting overall error rates, did not provide RT data for error trials or number of errors at each ISI, but it seems likely that such data might also have shown errors to vary inversely with ISI. This type of finding is of interest because it indicates that response conflict decreases as the response selection phase continues. This decrease in response conflict, suggested by both RT and error data in the present study, supports the continuous-strength Model B rather than the all-or-none Model A. Model B is also supported over Model A by the finding of Herman and Kantowitz (1969) that insertion of S_2 during the execution phase had no significant effect on MT; this shows that the ISI variable operates entirely prior to response execution, as indicated by Model B. Finally, the increment added to MT by the ISI variable is small, as might be expected from a stage fed by a degraded signal.

Theoretical Model B is testable since it implies that a weak competing response tendency inserted early in the selection phase can have as potent an effect as a strong competing response tendency inserted at the end of the selection phase. However, once the selection phase has ended, any competing response tendency will have minimal effect. The difficulty with this statement lies in fixing the end of the selection stage, for as long as the probability of response selection lies below 1.0, a sufficiently strong competing tendency may yet a fortiori replace it. Herman and Kantowitz (1969) avoided this problem by waiting at least 50

[4] The present study was concerned with the additivity of only two or three possible stages. It is likely that the influence of other variables not presently under investigation will lead to increases in the number of postulated stages.

msec. after RT to insert S_2. Additional research is needed to approach the end of the selection phase from below, e.g., with ISIs slightly less than RT.

Model B is also consistent with results reported by Lappin and Eriksen (1966). In their procedure, S_2 when presented was delayed 0, 12, 33, or 63 msec. The *S*s were instructed to respond to S_1 only if S_2 did not occur. As the ISI increased, *S*s became less able to inhibit their response to S_1. Thus once a potential response passes the vertical dashed line in Fig. 2, it cannot be recalled. Lappin and Eriksen also concluded that the motor portion of RT "seems to be an event with relatively constant duration, which is unmodifiable once initiated [p. 810]." This too is in accord with a one-stage execution phase.

REFERENCES

FITTS, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 1954, **47**, 381–391.

FITTS, P. M., & PETERSON, J. R. Information capacity of discrete motor responses. *Journal of Experimental Psychology*, 1964, **67**, 103–112.

HELSON, H., & STEGER, J. A. On the inhibitory effects of a second stimulus following the primary stimulus to react. *Journal of Experimental Psychology*, 1962, **64**, 201–205.

HERMAN, L. M. Effects of second signals on response time to first signals under certainty and uncertainty. *Journal of Experimental Psychology*, 1969, **80**, 106–112.

HERMAN, L. M., & ISRAEL, A. Decremental and facilitatory effects of second signals on response time to first signals under different levels of uncertainty. *Proceedings of the 75th Annual Convention of the American Psychological Association*, 1967, **2**, 27–28.

HERMAN, L. M., & KANTOWITZ, B. H. Effects of second signals occurring after response selection on responses to first signals. *Journal of Experimental Psychology*, 1969, **80**, 570–572.

HERMAN, L. M., & KANTOWITZ, B. H. The psychological refractory period effect: Only half the double stimulation story? *Psychological Bulletin*, in press.

HERMAN, L. M., & MCCAULEY, M. E. Delay in responding to the first stimulus in the "psychological refractory period" experiment: Comparisons with delay produced by a second stimulus not requiring a response. *Journal of Experimental Psychology*, 1969, **81**, 344–350.

KOPLIN, J. H., FOX, R., & DOZIER, R. A failure to replicate the inhibitory effects of a second stimulus following the primary stimulus to react. *Journal of Experimental Psychology*, 1966, **72**, 914–916.

LAPPIN, J. S., & ERIKSEN, C. W. Inhibition of a simple visual reaction time by a second stimulus: A failure to replicate. *Psychonomic Science*, 1964, **1**, 293–294.

LAPPIN, J. S., & ERIKSEN, C. W. Use of a delayed signal to stop a visual reaction time response. *Journal of Experimental Psychology*, 1966, **72**, 805–811.

STERNBERG, S. The discovery of processing stages: Extension of Donders' method. Paper presented at the Donders Centenary Symposium on reaction time, Instituut Voor Perceptie Onderzoek, Eindhoven, Netherlands, August 1968.

(Received March 17, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 353-359

AVERSION THRESHOLDS AND AVERSION DIFFERENCE LIMENS FOR WHITE LIGHT IN ALBINO AND HOODED RATS[1]

BYRON A. CAMPBELL[2] AND RITA B. MESSING

Princeton University

Preference thresholds for light vs. darkness and for one illumination of light over another were determined for albino and hooded rats, using a simple spatial preference test. Neither strain preferred any illumination level over total darkness, and the light aversion thresholds were the same for both strains. The magnitude of the aversion difference limen relative to illumination (Weber ratio) decreased drastically as illumination increased and then leveled off or increased somewhat at higher illuminations, with the Weber ratios of the hooded rats being about 1 log unit larger than those of the albinos. Comparison of the Weber ratios of aversiveness with human detection functions revealed that rats are quite sensitive to changes in illumination well within the photopic range.

For the rat, white light is thought to have both rewarding and aversive properties. Numerous investigators have shown that rats will perform an instrumental response to terminate bright light (e.g., Kaplan, Jackson, & Sparer, 1965; Keller, 1941), although the vigor of responding does not approach that produced by electric shock-induced pain or even mild degrees of hunger. Conversely, other investigators have shown that rats will press a bar to produce brief periods of light when the animal is housed in total darkness (cf. Lockard, 1963b, for a review).

Within this broad area of research, the major issue of current interest concerns the nature of the reinforcing effects of light onset. Two hypotheses have been advanced. One states that the reinforcing effect is due to the stimulus change or novelty produced by light onset (McCall, 1965; 1966), and the other states that some levels of light are inherently preferred over others and that animals will work to achieve that stimulus condition (Lockard, 1963a; 1964; 1966).

Part of this controversy stems, in our opinion, from an absence of definitive data on the basic preference-aversion function of rats for light. Two types of information are missing. First, what, if any, levels of light are preferred by rats over total darkness? Second, what magnitude of change in light intensity is necessary to produce preference or aversion from any given light intensity? This latter function is analogous to the difference limen or Weber ratio in psychophysics, only in this instance we are asking how much of a change is necessary to be rewarding or aversive. Accordingly, the purpose of the present research was to determine preference-aversion thresholds and aversion difference limens for white light with both albino (Sprague-Dawley) and hooded (Long Evans) rats.

This work can also be related to previous psychophysical analyses of electric shock (Campbell, 1955; Campbell & Masterson, 1969) and white noise (Campbell, 1968) in which aversion thresholds and aversion difference limens were obtained. Here it is of interest to compare the relative range over which the various kinds of stimuli are aversive and to compare the relative size (Weber ratios) of the aversion difference limens as a function of illumination. These data provide the investigator with common reference information regarding the reinforcing and punishing effects of stimulus changes in these modalities.

---

[1] This research was supported in part by Grants GB-5628 and BG-8046 from the National Science Foundation while the second author was a National Science Foundation predoctoral fellow. The authors wish to thank Stephie Orsi for her assistance in testing the animals and analyzing the data.

[2] Requests for reprints should be sent to Byron A. Campbell, Department of Psychology, Green Hall, Princeton University, Princeton, New Jersey 08540.

was fitted by the least-squares method. For the higher illuminations, the 75% and 25% preference points were interpolated, and the aversion difference limen (Δ I) was calculated by dividing the difference in illumination between the 75% and 25% points in half. At lower illuminations, comparisons lower than the standard were not run because the standard illuminations were very close to 0 ftc., making it impossible to choose comparison levels smaller than the standard by the same amount as those greater than the standard. When no 25% point was obtained, the aversion difference limen (Δ I) was taken as the difference between the illuminations producing 75% and 50% preference on the fitted function. From the aversion DLs, Weber ratios (Δ I/I, where I = standard intensity) were calculated. These ratios are shown in the left panel of Fig. 2. Here it is clearly evident that the Weber ratios for aversiveness decrease drastically as standard intensity increases and that the Weber ratios are approximately 1 log unit larger for the hooded rats than for the albino rats at all illumination levels. For the albino rat, it appears that the Weber ratio reaches a minimum around 11 ftc. and then begins to increase. This finding corresponds to the results of another experiment now in progress in our laboratory in which we find that the rate at which aversiveness of white light increases, when compared to the aversiveness of white noise or electric shock, drops off rapidly around 50 ftc. Thus it is not surprising that the Weber ratio should begin to increase at higher illuminations.

Figure 3 shows how the aversion DLs change over the course of the 100-min. test session. For the purposes of clarity, only four standard levels are shown for each strain. From Fig. 3 it is evident that asymptotic preference for the less intense light is

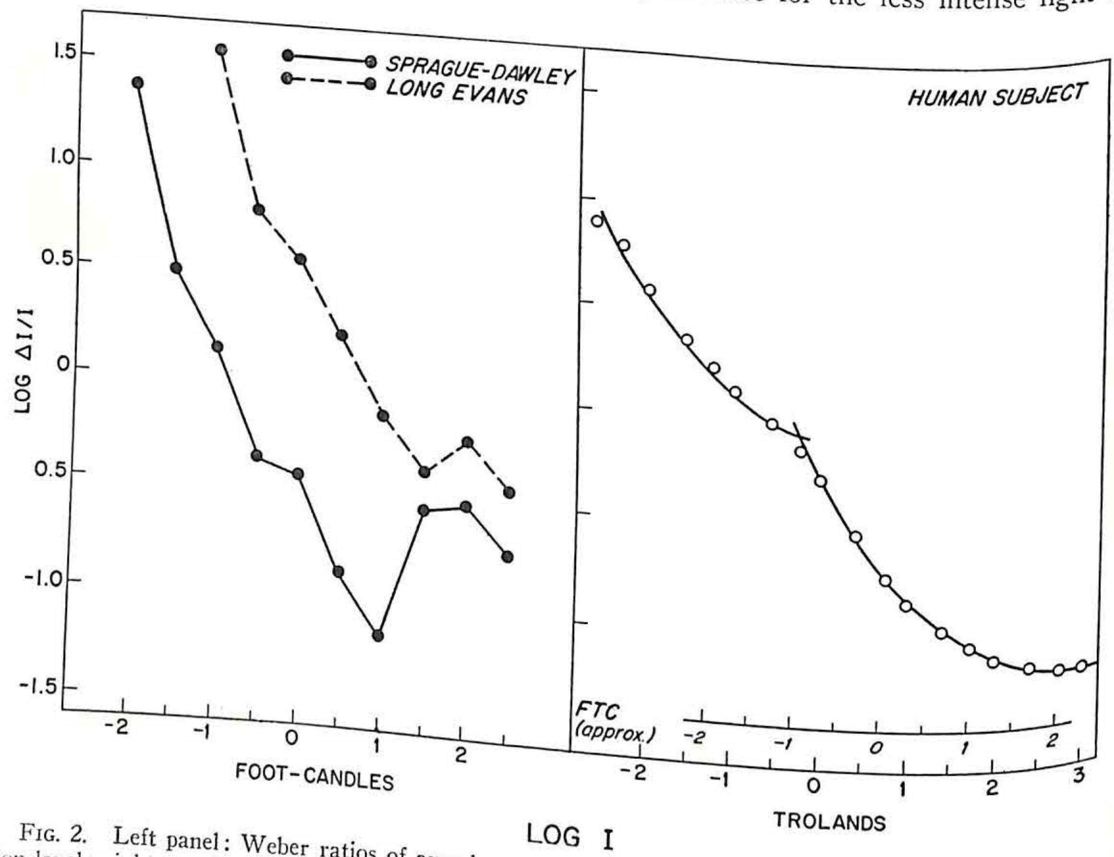


FIG. 2. Left panel: Weber ratios of aversiveness for albino and hooded rats as a function of illumination level; right panel: Weber ratios (detection) for a typical human *S* (from Hecht, Peskin, & Patt, 1938).

just barely reached, if it is reached at all, within the 100-min. session. Quite possibly, longer test sessions would have produced even sharper preference functions.

DISCUSSION

Perhaps the most striking feature of the present study was the failure to find any intensity of light that was preferred by rats of either strain significantly more than total darkness. Even the hooded rat, presumably a more visual animal, showed no absolute preference for light under the present experimental conditions. These results are in contrast with some previous findings, particularly those of Lockard (1963a; 1963b; 1964; 1966), who found that rats showed a preference for low levels of illumination when onset and offset of light was controlled by two separate levers. This preference for light, however, decreased over sessions, and it is our view that a simple spatial preference apparatus that does not require the acquisition of a complex instrumental response allows the habituation to novelty produced by light onset and offset to proceed extremely rapidly, thus permitting the rats' natural aversion to light to manifest itself in the course of a single 100-min. session. In the only other experiment using simple spatial preference as the criterion response, Hanson (1951) also found that rats showed no preference for light of any intensity.

In the present study, *S*s were not dark-adapted prior to testing, and therefore it is possible that the aversion thresholds might be even lower in dark-adapted animals. Dark adaptation in the rat takes many hours (Dodt & Echte, 1961) so that even at the end of the 100-min. session, *S*s were not fully adapted. It is possible to argue that previous exposure to moderate illumination levels in the home cage (about .3–3 ftc.) satiated "preference" for light, but the present investigators believe this explanation implausible in view of Lockard's (1963a; 1964) finding that rearing luminance varies directly with preference for test luminance.

It should also be emphasized that the low aversion thresholds do not imply that rats find light aversive under all circumstances and thus never venture into light places when they can be avoided. Gibson (1966) points out the need to regard different perceptual systems as interrelated and contrasts dynamic perception with the activation of essentially passive receptors as studied in the laboratory. While the environment in the present experiment does not make this a study of passive receptors, neither does it constiute a natural situation for the animal in which many types of stimulation impinge and in which there is a great deal of information to be gathered from visual stimuli that can lead to adaptive responses.

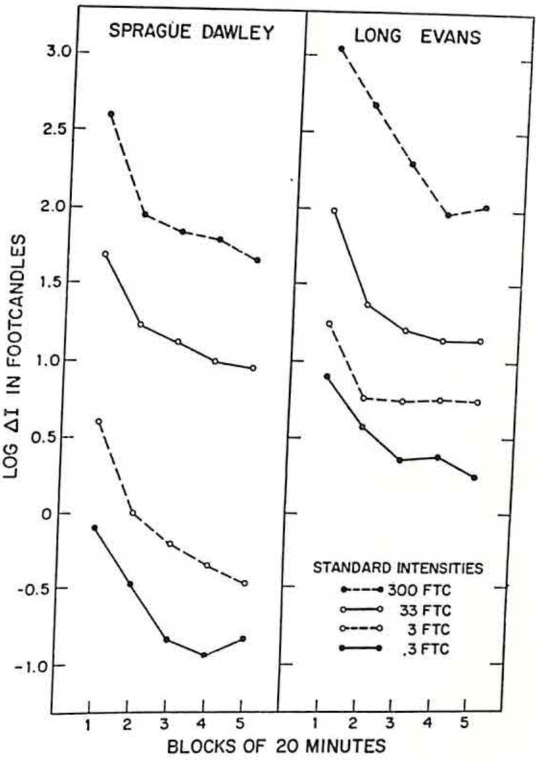


FIG. 3. Aversion difference limens for albino and hooded rats during successive 20-min. periods of the 100-min. session at four representative standard illumination levels.

Comparison of the number of "just aversive differences" (the equivalent of the JND or DL in human psychophysics) for light, shock (Campbell & Masterson, 1969), and white noise (Campbell, 1968) clearly shows that the number of "just aversive differences" (JADs) between two levels of aversive stimulation has little, if anything, to do with the aversiveness of one stimulus compared to another. Unpublished data from our laboratory (Messing, 1969) show that .5 ftc. is equal in aversiveness to .047 ma. of electric shock and 87 db. of noise, while at the high end of the intensity continuum, 50 ftc. equals .097 ma. of shock and 115 db. of noise. Between these equally aversive stimulus intensities (for each modality) there are approximately 7 JADs for shock (Campbell & Masterson, 1969), 24 JADs for noise

(Campbell, 1968), and 32 for light. If the JAD reflected an equal increment in aversiveness, then the cumulated number of JADs between equal aversion contours should be the same, regardless of sensory modality. Thus it is evident that the JAD cannot be used as the unit of aversiveness, just as the JND cannot be used as the unit of stimulus magnitude in human psychophysics (Stevens, 1957).

Similarly, it is possible to compare the Weber ratios for aversiveness obtained in the present study with the Weber ratios obtained in human psychophysics. For this purpose the right panel of Fig. 2 shows a typical function for a single human *S* (Hecht, Peskin, & Patt, 1938). The major difference between the two functions is the presence of a distinct rod-cone break in the human function at about −.3 log trolands. In addition, the Weber ratios are smaller for humans than the aversion DLs at practically all illuminations. This is not surprising since one would not typically expect the aversion DL to be as small as the just-detectable difference. Moreover, the experimental conditions of Hecht et al. were much more conducive to obtaining smaller limens. If aversiveness could be measured under circumstances comparable to those used in human research, e.g., fixed focal point, contrast between two illuminated concentric fields, prior dark adaptation, etc., it is likely that the aversion DL would be much smaller.

Nevertheless, comparison of the two curves suggests that both the hooded and albino rats are able to distinguish smaller differences in illumination than can be detected by the rods of the human retina, although in humans increasing sensitivity of the rods at higher illuminations may be masked by the much greater sensitivity of the cones. However, a study on a human *S* with a nearly all rod retina (Fuortes, Gunkel, & Rushton, 1961) indicated that the relative sensitivity of the rods does not increase appreciably beyond the point where the rod-cone break is typically found. Moreover, sensitivity decreased sharply between 2–3 log trolands, and *S* was unable to distinguish illumination changes above 3 log trolands. Albino and hooded rats were still quite sensitive to illumination change at the highest illumination the present investigators were able to run (equivalent to approximately 2.6 trolands in the Fuortes paper and about 3.1 log trolands in the Hecht et al., 1938, paper).

In the overview it appears that the rat responds to changes in illumination at least 2.5 log ftc. above the rod-cone break in humans (about −1 log ftc. or −.3 log trolands in the Hecht et al., 1938, paper) and therefore may have a degenerate photopic receptor system. While this may be surprising to those who hold the view that rats have all-rod retinas, it makes sense from an evolutionary viewpoint (Polyak, 1957). The early mammals undoubtedly evolved from keen-sighted diurnal reptiles, and while many of these animals became nocturnal in order to avoid extinction by reptilian predators, it is not surprising that some vestige of the diurnal retina may remain in the rat.

REFERENCES

CAMPBELL, B. A. The fractional reduction in noxious stimulation required to produce "just noticeable learning." *Journal of Comparative and Physiological Psychology*, 1955, **48**, 141-148.

CAMPBELL, B. A. Interaction of aversive stimuli: Summation or inhibition? *Journal of Experimental Psychology*, 1968, **78**, 181–190.

CAMPBELL, B. A., & MASTERSON, F. A. Psychophysics of punishment. In B. A. Campbell & R. M. Church (Eds.), *Punishment*. New York: Appleton-Century-Crofts, 1969.

DODT, E., & ECHTE, K. Dark and light adaptation in pigmented and white rat as measured by electroretinogram threshold. *Journal of Neurophysiology*, 1961, **24**, 427–445.

FOURTES, M. G. F., GUNKEL, R. D., & RUSHTON, W. A. H. Increment thresholds in a subject deficient in cone vision. *Journal of Physiology*, 1961, **156**, 179–192.

GIBSON, J. J. *The senses considered as perceptual systems*. Boston: Houghton-Mifflin, 1966.

HANSON, L. F. Light avoidance and light aversion in the albino rat as a function of intensity of illumination. Unpublished doctoral dissertation, Columbia University, 1951.

HECHT, S., PESKIN, J. C., & PATT, M. Intensity discrimination in the human eye. II: The relation between ΔI/I and intensity for different parts of the spectrum. *Journal of General Physiology*, 1938, **22**, 7–19.

KAPLAN, M., JACKSON, B., & SPARER, R. Escape behavior under continuous reinforcement as a function of aversive light intensity. *Journal of the Experimental Analysis of Behavior*, 1965, **8**, 321–323.

KELLER, F. S. Light aversion in the white rat. *Psychological Record*, 1941, **4**, 235–240.

LOCKARD, R. B. Self-regulated exposure to light by albino rats as a function of rearing luminance and test luminance. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 558-564.

LOCKARD, R. B. Some effects of light upon the behavior of rodents. *Psychological Bulletin*, 1963, **60**, 509–529.

LOCKARD, R. B. Self-regulated exposure to light by pigmented and albino littermates. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 231–236.

LOCKARD, R. B. Several tests of stimulus-change and preference theory in relation to light-controlled behavior of rats. *Journal of Comparative and Physiological Psychology,* 1966, **62**, 415–426.

McCALL, R. B. Stimulus change in light-controlled bar pressing. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 258–262.

McCALL, R. B. Initial-consequent change surface in light-contingent bar pressing. *Journal of Comparative and Physiological Psychology,* 1966, **62**, 35–42.

MESSING, R. B. Experiments in animal psychophysics. Unpublished doctoral dissertation, Princeton University, 1969.

POLYAK, S. *The vertebrate visual system.* Chicago: University of Chicago Press, 1957.

STEVENS, S. S. On the psychophysical law. *Psychological Review,* 1957, **64**, 153–181.

(Received March 3, 1969)

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Semantic Generalization without Prior Conditioning: I. Maltzman,* B. Langdon, and D. Feeney: Department of Psychology, University of California, Los Angeles, California 90024.

Preference for Immediate versus Delayed Noxious Stimulation and the Concomitant GSR: Irving Maltzman* and Craig Wolff: Department of Psychology, University of California, Los Angeles, California 90024.

Naming and Other Methods of Decoding Visual Information: Richard L. Taylor* and Stephen Reilly: Veterans Administration Hospital, Irving and University Place, Syracuse, New York 13210.

Response Elimination in Noncorrection Paired-Associates Learning: Ludwig Mosberg*: Southwest Regional Laboratory, 11300 La Cienega Boulevard, Inglewood, California 90304.

Intradimensional Variability and Concept Identification: R. C. Haygood,* T. L. Harbert, and J. A. Omlor: Department of Psychology, Anderson Hall, Kansas State University, Manhattan, Kansas 66502.

Information Processing Rates and Task Complexity: Richard W. Olshavsky* and Lee W. Gregg: School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332.

Visual Masking in Multielement Displays: Charles W Eriksen* and John Rohrbaugh: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

Divided Attention: A Vehicle for Monitoring Memory Processes: W. A. Johnston,* S. N. Greenberg, R. P. Fisher, and D. W. Martin: Human Performance Center, Ohio State University, 404-B West Seventeenth Avenue, Columbus, Ohio 43210.

Operant Suppression of Electrodermal Response Rate as a Function of Punishment Schedule: A. Crider,* G. E. Schwartz, and D. Shapiro: Department of Psychology, Williams College, Williamstown, Massachusetts 01267.

Analysis of the Differences between Free and Serial Recall: Gail A. Bruder*: Department of Psychology, State University of New York at Buffalo, Buffalo, New York 14214.

Variation in Reports of Covert Rehearsal and in STM Produced by Differential Payoff: W. E. Montague,* W. A. Hillix, H. O. Kiess, and R. Harris: Department of Psychology, University of Illinois, Urbana, Illinois 61801.

Semantic Generalization: IAR Locus and Instructions: Phebe Cramer*: Institute of Human Learning, University of California, Berkeley, California 94720.

Dimensional Dominance and Continuity Theory: T. S. Kendler,* B. H. Basden, and J. B. Bruckner: Department of Psychology, University of California, Santa Barbara, California 93106.

Recognition Reaction Time for Digits in Consecutive and Nonconsecutive Memorized Sets: Donald V. DeRosa and Robert E. Morin*: Department of Psychology, Kent State University, Kent, Ohio 44240.

---

* Asterisk indicates author for whom address is given.

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 360–365

EFFECT OF SEMANTIC REDUNDANCY ON CHILDREN'S IDENTIFICATION OF VERBAL CONCEPTS[1]

FRANCIS J. DI VESTA[2] AND GARY M. INGERSOLL

Pennsylvania State University

The effects of cue-oriented instructions, redundancy, and attribute saliency on concept identification in a verbal task were investigated. Attributes were defined as the three salient dimensions of semantic meaning. The *Ss* were 130 children in Grades 4–6 inclusive. Increased cue saliency, provided through instructions limiting the number of attributes (semantic dimensions) to be identified, facilitated performance. The effects of redundancy and attribute saliency were found within the cue-oriented groups. Concepts were found to be identified more easily when lists comprised of words with three relevant attributes were used than when lists comprised of words with one or two relevant attributes were used. In addition, concept identification was facilitated to a greater extent in lists where the evaluative dimension was relevant than in lists where it was irrelevant. Saliency of cues was interpreted as having its effect on what is discriminated; the degree of complexity was interpreted as having its effects on making the concept to be identified less ambiguous.

Considerable evidence has accumulated indicating that concept identification is affected by the number of relevant and irrelevant cues (Archer, Bourne, & Brown, 1955; Bulgarella & Archer, 1962). The addition of irrelevant cues tends to increase the complexity of the problem with a corresponding increase in the difficulty of concept identification (Archer et al., 1955). Conversely, increases in the number of redundant dimensions among those which are relevant have a facilitating effect on concept identification. Thus it can be expected that the facilitating effects of redundancy will partially offset the detrimental effects of irrelevant attributes on performance (Bourne & Haygood, 1959; 1961). These effects apply to children's learning as well as to adults' learning (Osler & Kofsky, 1965).

In the typical concept identification study, stimulus attributes are often depicted pictorially and the stimulus cues are considered to be of approximately equal strength. However, it is reasonable to hypothesize that if the relevant attributes were of differing strength, there would be corresponding differences in performance. Of particular importance for the present study is Archer's (1962) finding that when the saliency (obviousness) of a relevant dimension is increased, performance improves; when the saliency of an irrelevant dimension is increased, performance gets worse, presumably because *S*'s attention is directed to that attribute. The important feature of Archer's results for the present study is that some dimensions are more salient than others, and this can be manipulated. Measures of semantic meaning would provide one opportunity for examining the effects of such variables on concept identification.

Some recent studies indicate the feasibility of using dimensional relations of semantic structure in concept formation studies (e.g., Di Vesta & Walls, 1969b; Haygood, 1966; Taylor & Haygood, 1968) based on the three factors of evaluation, potency, and activity typically extracted in factor analytic studies. These dimensions account for the major portion of the variance defining semantic space, though in diminishing amounts (Di

---

[1] This research was supported by Grant HD-02025 from the National Institute of Child Health and Human Development, United States Public Health Service. The authors gratefully acknowledge the cordial cooperation of Ott B. Kantner, Director of Elementary Education in the Bellefonte Area School District, John Roberts, Principal of the Bellefonte Elementary School, the teachers, and their pupils.

[2] Requests for reprints should be sent to Francis J. Di Vesta, Department of Educational Psychology, 311 Education and Psychology Center II, Pennsylvania State University, University Park, Pennsylvania 16802.

Vesta & Walls, 1969b; Osgood, Suci, & Tannenbaum, 1957), and are considered as attributes in concept learning studies. If the variance accounted for by each dimension (attribute) reflects the degree of saliency in other settings, one would expect performance to be related to the relevant attribute. Thus, the evaluative dimension, typically extracted as the most salient dimension by factor analyses, should display comparable strength in a concept identification study in which semantic dimensions define the salient attributes.

Finally, semantic meanings are but a few of the very many different meanings or attributes that might be associated with a set of verbal stimuli. Accordingly, it would be expected that a condition in which the possible relevant dimensions were delineated by instruction, or other device, would induce better performance than one in which the possible alternatives were not defined. Simply, an *S* who is not given the limiting instruction has a much wider range of features from which he must choose the relevant attributes than does an *S* who is provided such instructions.

METHOD

Experimental design.—One of a series of eight lists, with 40 words in each, was randomly assigned for presentation to an *S* in a concept identification task. The *S*'s task was to determine the rule by which each word in the list could be classified into a dichotomous X-not-X categorization. The eight lists were varied according to the number and kind of relevant semantic attributes represented in the stimuli. The relevant attributes were defined as the three salient dimensions of semantic space. The selection of words was made from norms based on the ratings by fifth-grade children (Di Vesta & Walls, 1969b). The *S*s were further divided into two instructional conditions: In one treatment, the three dimensions to be used were delineated for *S* prior to his participation in the task; in the other, the instructions simply indicated that classification was to be based on meaning. There were 16 groups in all, resulting in a 2 × 8 factorial analysis of variance.

Subjects.—The *S*s were 130 fourth-, fifth-, and sixth-grade children. Each *S* was assigned, at random, to 1 of the 16 treatments on arrival at the experimental setting. There were 2 *S*s dropped from the experiment because of extreme anxiety.

Stimulus materials.—The relevant and irrelevant features of the stimulus words were defined in terms of the three salient dimensions of semantic space. In all, eight lists were constructed from normative data for fifth-grade children (Di Vesta & Walls, 1969a): Three lists had one relevant feature; in each list the stimulus words were drawn from the high- and low-polarization end of either the evaluative (E), potency (P), or activity (A) dimension of semantic space as defined by Osgood et al. (1957). Thus, e.g., the words for List A were selected so that their rating on the A dimension was relevant, whereas their ratings on the E and P dimensions were not. A set of three more lists was constructed by selecting words in which two features were simultaneously relevant in each list, thereby yielding combinations of E and P (EP), E and A (EA), or P and A (PA) as relevant attributes. It should be noted that simultaneously relevant means *redundant* and does *not* refer to such combinations of attributes as are found in conjunctive or disjunctive concepts. Thus, e.g., ratings on both the E and P dimensions were relevant and the rating on the A dimension was irrelevant for words in the EP list. Another list was comprised of words for which the ratings on all three dimensions were relevant (List EPA). A final control (C) list consisted of words in which there was no consistent relevant dimension, thereby paralleling the use of the verbal discrimination task as described by Underwood (1966) and providing a base line in which learning-to-learn, warm-up, and discriminative processes were considered.

The words were selected on the basis of factor scores calculated according to the procedure described by Glass and Maguire (1966). This procedure is based on all scales employed in the factor analysis of ratings and yields factor scores that are orthogonal to one another (Di Vesta & Walls, 1969a). Factor scores of all X words were polarized at one end of the relevant dimension(s), and O (not X) words were polarized at the other end of the relevant dimension(s); the factor scores of X and O words did not overlap on the relevant dimensions, and their variances were small. The constraint that words be rated on the extremes for one or more factors and within the neutral range on the remaining factor(s) was impossible to impose for all lists. Accordingly, selection was based on the rule that factor scores of the stimulus words on *irrelevant* dimensions should vary through the entire range of scores; thus the mean factor scores of words for irrelevant dimensions did not differ significantly, and their variances were high. A complete listing of means and variances of factor scores on the three dimensions for all eight lists can be found in Table 1.

Procedure.—Each *S* was administered the task individually. Those *S*s in the non-cue-oriented (NCO) instructional condition were provided standard concept identification instructions, but were not told the specific dimensions on which the stimuli were varied, other than the fact that meaning was important. Those *S*s in the cue-oriented (CO) instructional condition received, in addition

TABLE 1

MEANS AND STANDARD DEVIATIONS OF FACTOR SCORES FOR X AND O WORDS IN ALL LISTS

List	Dimension					
	Evaluative		Potency		Activity	
	M	*SD*	*M*	*SD*	*M*	*SD*
E						
X	1.64	.31	3.70	1.32	4.01	.95
O	4.96	.58	3.25	1.10	3.53	1.18
P						
X	3.27	.95	1.32	.48	3.44	1.07
O	2.74	1.34	4.84	.59	3.22	1.32
A						
X	2.85	1.60	3.07	1.32	1.56	.32
O	2.72	.95	2.60	1.24	5.19	.49
EP						
X	1.72	.35	1.78	.54	4.17	.88
O	4.55	.54	3.91	.56	3.72	1.05
PA						
X	3.46	1.74	1.74	.47	2.05	.37
O	3.10	1.29	4.21	.59	4.48	.61
EA						
X	1.79	.36	3.29	.98	1.82	.38
O	4.77	.59	3.01	1.03	4.34	.32
EPA						
X	1.66	.33	2.20	.46	2.32	.43
O	4.57	.68	3.67	.46	4.07	.48
C						
X	2.69	1.13	3.17	1.39	3.66	1.26
O	2.59	1.18	2.85	1.20	3.80	1.04

Note.—These factor scores are not standardized. Thus the neutral point does not necessarily correspond to the semantic differential scale midpoint of 4.00.

to the standard concept identification instructions, a description of the three semantic meaning dimensions as a way in which meaning of words might vary. (Groups are identified by symbol and number to indicate instructional condition and number of relevant stimuli; thus, CO-1 indicates cue-oriented instructions with the concept based on one relevant attribute.)

Each *S* was presented four random presentations of one of the experimental lists of 40 words by means of a Kodak Carousel projector. Each word was rear-projected onto a translucent plastic screen at a rate, paced to the abilities of *S*, of one slide about every 4–5 sec. The *S* read the word aloud and then indicated whether it belonged in the X or O (not X) category. After *S* responded, *E* informed him of the correct response. This procedure was followed for three complete presentations of the list. The complete experiment took approximately 20 min.

RESULTS

The data for the first presentation of the list were analyzed by a mixed analysis of variance in which the eight lists were orthogonally crossed with the two instructional treatments. The total number of words was divided into four 10-trial segments. This analysis yielded significant effects due to lists, $F\ (7, 112) = 4.09$, $p < .01$, to instructions, $F\ (1, 112) = 23.84$, $p < .01$, and to blocks of trials, $F\ (3, 336) = 3.87$, $p < .05$. None of the interactions was significant, though the Instructions × Trials interactions yielded $F\ (3, 336) = 2.48$, $p < .10$. The mean number of correct responses was directly related to the number of relevant attributes within the list. Significantly more correct responses were made under CO than under NCO instructions. The relevant data for this analysis are presented graphically in Fig. 1 with additional data to permit comparison with that presented in Fig. 2.

Tests of simple effects were performed using Dunn's technique, which yields *d* values for each comparison examined (Kirk, 1968), in order to determine the extent to which the specific hypotheses were supported. Although the means for different lists occurred in the order predicted, the difference between lists with one and two

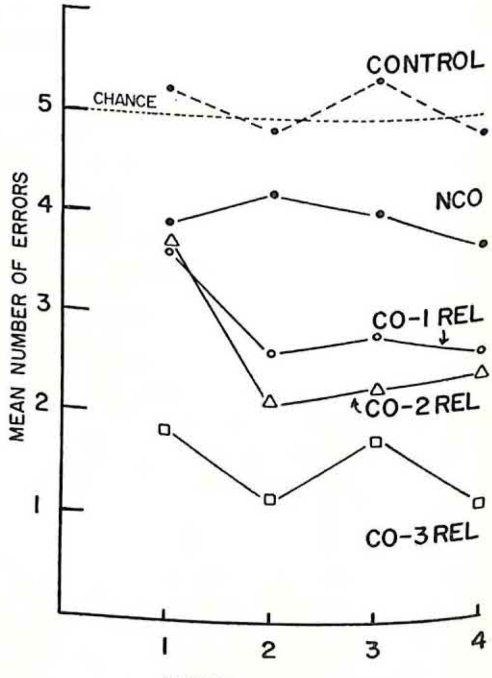


FIG. 1. Mean number of errors in concept identification as functions of cue orientation (CO and NCO) and the number of relevant (REL) attributes.

redundant dimensions in the CO conditions was not significant, $d = 1.61$, $p > .05$. However, there were significant differences between group means for lists with one and three redundant dimensions, $d = 5.10$, $p < .01$, and for lists with two or three redundant dimensions, $d = 3.98$, $p < .01$. A comparison of means for *S*s learning lists in which the evaluative dimension was relevant and lists in which it was irrelevant yielded $d = 4.92$, $p < .01$. A post hoc comparison, via Scheffé's test, of performance on lists in which activity was the relevant dimension with lists in which it was irrelevant yielded $S = 5.88$, $p < .01$. Parallel analyses of simple effects in the NCO treatments did not yield significant results. All experimental treatments produced significantly ($p < .001$) more correct responses than the control.

All experimental groups not only made more correct responses than did the control group on the first presentation, but their performance on subsequent presentations improved more rapidly than that of the control group, as can be seen in Fig. 2. However, inspection of the performance curve for Group EPA-CO indicates little improvement over presentations, primarily because they identified the concept so rapidly.

The analysis of variance for data summarized for all three presentations in Fig. 2 was based on dividing each series into 20-trial blocks. The results were essentially the same as those based on the analysis of data from the first presentation of the list. As hypothesized, the effect due to kind of instruction was significant, $F(1, 112) = 22.59$, $p < .001$. With the exception of the two control groups, overall performance by *S*s in all groups was facilitated more by the CO than by the NCO instructions. The effect due to trials yielded $F(1, 112) = 2.48$, $p < .20$. The Kind of Instruction × Trials interaction yielded $F(1, 112) = 2.80$, $p < .10$. The Lists × Kind of Instructions interaction was not significant, $F(7, 112) = 1.89$, $p > .05$. By comparing the results of this analysis with those of the first analysis, it can be seen that the major effects of the variables occur in the initial presentation of the list.

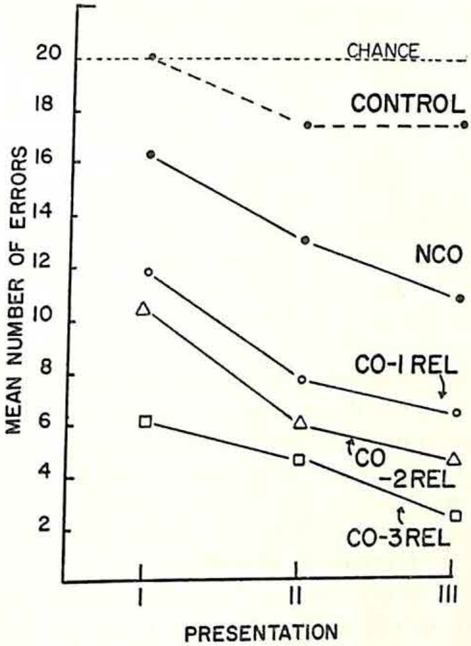


Fig. 2. Mean number of errors in concept identification over three complete presentations of the 40-word list as functions of cue orientation and the number of relevant (REL) attributes.

In order to determine the extent to which the specific hypotheses were supported over all presentations of the list, tests of simple effects were performed using Dunn's technique. This analysis again yielded similar results to those reported for the analysis of the data for the first presentation. The means for lists with one or two redundant features occurred in the order predicted. However, the difference was not significant, $d = 2.13$, whereas $d\ (.05) = 2.54$. Significantly more correct responses were made by *S*s learning the list with three redundant dimensions than were made by *S*s learning the lists with either one, $d = 7.00$, $p < .01$, or two, $d = 5.48$, $p < .01$, redundant dimensions. The trend in performance was clearly related to the number of redundant dimensions: the greatest number of correct responses was evoked by *S*s employing the list with three relevant attributes; fewer correct responses were made by *S*s employing the lists with two relevant attributes; and the least number of correct responses was made by *S*s employing the list with only one attribute. Within the level of one redundant feature, more errors were made in the combined results from Lists P and A compared to the results from List E, $d = 6.06$, $p < .01$; with two redundant attributes, more errors were made with List PA compared to the combined results of Lists EP and EA, $d = 2.92$, $p < .05$. In addition, those three lists containing the A dimension (Lists EA, PA, and A) elicited fewer errors than those not containing the A dimension. This difference was significant via Scheffé's test, $S = 4.70$, $p < .01$. These results clearly reflect the differential effects of the kind of attributes and the saliency of the evaluative dimension as well as the effects of the number of redundant attributes. Similar analyses of data for the NCO treatments yielded no significant differences related to the number of redundant dimensions. Finally, *S*s in the control groups produced significantly more errors than did those *S*s in all other treatments combined, $d = 6.57$, $p < .01$.

DISCUSSION

The results of the present study clearly demonstrate differences in performance related to the kinds of information *S* is provided about the attributes in a concept identification task, to the number of relevant attributes on which formation of the concept depends, and to the saliency of the attributes of the concept. These effects were demonstrated with verbal (rather than pictorial) stimuli in which the attributes were defined in terms of their semantic meanings (dimensional relationships) and with children (rather than adults) as *S*s.

The control group's performance is assumed to reflect the degree of recall of the specific status of given stimuli. Thus on the initial presentation of the series, the group performed at chance level. This result would be expected, of course, since all stimuli were different and only one of two responses (X or O) had to be made. On subsequent trials, there was some slight improvement in performance undoubtedly due to the associations made in the first presentation of the list. Compared to the control group, the performance of all other groups was facilitated on all blocks of trials. The increment in performance during the first presentation was considered to be due to the acquisition of the concept since all stimuli were unique in this phase. The abstraction of the correct dimensional attribute(s) was the only basis on which the performance of any experimental group could exceed that of the control group. This is not to suggest that all improvement beyond the first 40-trial block is rote learning. In the more difficult conditions (especially CO-1 and NCO), *S*s may not be acquiring the concept until late in the first 40 trials. Accordingly, its advantage does not become apparent until the later presentation.

On the basis of the previous rationale, it was reasoned that the NCO groups acquired some basis for identifying the concept relative to the control group, but it was not sufficiently clear to identify all concept instances. There was some evidence that *S*s in these groups may have employed an attribute other than the one defined by *E*, but which simultaneously resulted in their being correct more often than by chance. Specifically, the activity dimension is such that words selected according to degree of activity are often exemplars of animate and inanimate objects. Accordingly, *S*s' attention to the activity dimension was preempted by what may be the more salient animate–inanimate classification. Though these two dimensions are related, they are not perfectly correlated; errors were made by these *S*s, and redundancy failed to improve their performance. In the CO conditions, the range of possible attributes that needed to be considered by *S*s was limited. The result was a reduction in the complexity of the concept with a corresponding improvement in their performance.

The major concern in the present study, however, was the effect on concept complexity as it is affected by manipulations of the number of relevant dimensions. When relevant redundancy was introduced, it aided performance, as in the Bourne and Haygood (1959) study. This is especially true in the groups oriented to the cues of important attributes where, in effect, the number of attributes from which *S*s were to make their selection when identifying the concept was clearly delimited. Thus where one dimension was relevant, *S*'s task was to determine which attribute was relevant, then to determine whether positive or negative polarization was associated with the X or *O* categories, and, finally, to determine the rule for combining attributes. Where three attributes were relevant, the initial process is eliminated. This observation also indicates that a difficulty with the present design is that the number of irrelevant dimensions is being decreased as the number of redundant dimensions is increased. However, such confounding can never be eliminated completely since it may be difficult to specify exactly which dimensions are present for a given *S*.

Similar effects were not particularly apparent in the NCO groups presumably because the complexity of the task was such that it overrode the effect of redundancy. Since *S*s in all NCO conditions performed very much alike, the evidence strongly suggests that concepts may have been formed mainly on the polarization of words as an attribute rather than on the more specific dimensional relationships among words.

Finally, the effect of dimensional attributes of verbal materials on concept complexity was further corroborated in this study. The hypothesis that concepts based on the evaluative dimension would be learned more easily than those based on either the potency or activity dimension was supported. These results reflect the saliency of the evaluative dimension. In factor analyses of ratings of words obtained from both children (Di Vesta & Walls, 1969a) and adults (Gerow & Pollio, 1965; Osgood et al., 1957), the evaluative factor typically emerges as the most dominant one. The relative saliency of the activity and potency dimensions, from the studies on which the aforementioned finding emerged, is less clear, except that neither is as dominant as the evaluative dimension, a finding also reflected in the present results.

REFERENCES

ARCHER, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. *Journal of Experimental Psychology*, 1962, **63**, 616–620.

ARCHER, E. J., BOURNE, L. E., JR., & BROWN, F. G. Concept identification as a function of irrelevant information and instructions. *Journal of Experimental Psychology*, 1955, **49**, 153–164.

BOURNE, L. E., JR., & HAYGOOD, R. C. The role of stimulus redundancy in concept identification. *Journal of Experimental Psychology*, 1959, **58**, 232–238.

BOURNE, L. E., JR., & HAYGOOD, R. C. Effect of redundant relevant information upon the identification of concepts. *Journal of Experimental Psychology*, 1961, **61**, 259–260.

BULGARELLA, R. G., & ARCHER, E. J. Concept identification as a function of amount of relevant and irrelevant information. *Journal of Experimental Psychology*, 1962, **63**, 254–257.

DI VESTA, F. J., & WALLS, R. T. A factor analysis of the semantic attributes of 485 words and some relationships to the conceptual behavior of fifth-grade children. (Tech. Rep. No. 15, Contract No. HD-02025) University Park: Pennsylvania State University, 1969. (a)

DI VESTA, F. J., & WALLS, R. T. Effects of rule and attribute identification on children's attainment of disjunctive and conjunctive concepts. *Journal of Experimental Psychology*, 1969, **80**, 498–504. (b)

GEROW, J. G., & POLLIO, H. R. Word association, frequency of occurrence, and semantic differential norms for 360 words. (Tech. Rep. No. 1, Contract No. MH-08903) Knoxville: University of Tennessee, 1965.

GLASS, G. V., & MAGUIRE, T. O. Abuses of factor scores. *American Educational Research Journal*, 1966, **3**, 297–304.

HAYGOOD, R. C. The use of semantic differential dimensions in concept learning. *Psychonomic Science*, 1966, **5**, 305–306.

KIRK, R. E. *Experimental design procedures for the behavioral sciences.* Belmont, Calif.: Brooks/Cole, 1968.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

OSLER, S. F., & KOFSKY, E. Stimulus uncertainty as a variable in the development of conceptual ability. *Journal of Experimental Child Psychology*, 1965, **2**, 264–279.

TAYLOR, C. L., & HAYGOOD, R. C. Effects of degree of category separation on semantic concept identification. *Journal of Experimental Psychology*, 1968, **76**, 356–359.

UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1966.

(Received March 17, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 366–376

EYE TRACKING OF SELF-MOVED TARGETS: THE ROLE OF EFFERENCE[1]

MARTIN J. STEINBACH[2]

York University

Simultaneous monitoring of eye and hand positions shows that an *S* can visually track a target more accurately if moved by him rather than if moved independently of his control, indicating that the oculomotor system has access to the efference (outflow) that produces the target movements. In tracking, most of the useful information in the efferent signal appears to be dynamic and thus concerned with the self-moved target's absolute velocity and acceleration changes. The *S* was able to track smoothly an intermittently appearing target (which of itself gave no information about its motion) when the target movement was related to *S*'s passively moved arm, indicating a proprioceptive input from the arm of use to the pursuit eye movement system.

How can an organism distinguish sensory input resulting from its own movements from sensory events that occur independently of its movements? Both Holst (1954; Holst & Mittelstaedt, 1950) and, independently, Sperry (1950) suggested that the organism must take into account the movement-produced sensory change (reafference) by comparing it to a secondary motor outflow ("efferenzkopie" or "corollary discharge") that was issued along with the efferent signal that produced the movement. This comparison, then, allowed for the determination of whether the sensory stimulation was movement-produced or externally produced (exafferent). Teuber (1966) suggested that the central nervous system defines whether a movement is voluntary (self-produced) or involuntary (reflexive) by the presence or absence of corollary discharge, and he speculated that the origin of corollary discharge may be found in the frontal lobes (Teuber, 1964).

The reality of a monitored or secondary efference has been suggested by the results of several contemporary investigations. Held and his colleagues (e.g., Held & Freedman, 1963; Held & Hein, 1958) have shown that full adaptation to rearranged visual inputs (as when displacing prisms are worn over the eyes) depends to a great extent on having *S* move himself actively through his environment. If *S*'s movements are passive, i.e., not self-produced, little or no adaptation occurs. Held argued that an intact sensorimotor feedback loop, one complete with motor outflow and subsequent visual input, is necessary for the adaptive process; because there is no outflow in the passive condition, the process is restricted and adaptation is minimal.

More recently, Steinbach and Held (1968) showed that the oculomotor system could use outflow information issued to move the arm. In their experiment, *S* visually tracked a target mounted on his fingertip. The tracking performance was better—the eye lagged behind the target less and made fewer corrective saccades—when the target motion was self-produced (*S* moved his own arm) than when he relaxed his arm and allowed *E* to move it in a similar manner.

In the experiment of Steinbach and Held (1968), the tracking performance could have been better in the active movement condition because the oculomotor system could anticipate either the future positions of the target or its motions. Which information was being supplied in the outflow? The experiments to be reported here confirm and extend the earlier findings and indicate that motion, and more specifically, velocity and acceleration components are the relevant informational sources carried in the monitored efferent signal.

---

[1] This article is based on a dissertation submitted to the Massachusetts Institute of Technology for the PhD degree. The exceptionally helpful guidance and suggestions of Richard Held, Hans-Lukas Teuber, Whitman A. Richards, and Gerald Schneider are gratefully acknowledged. The research was supported by National Aeronautics and Space Administration Grant NsG 496 (to Teuber), National Institute of Mental Health Grant MH-07642 (to Held), and a National Defense Education Act, Title IV predoctoral fellowship held by the author.

[2] Requests for reprints should be sent to Martin J. Steinbach, Department of Psychology, York University, 4700 Keele Street, Downsview, Ontario, Canada.

EXPERIMENT I

Experiment I is a replication of Steinbach and Held (1968) under conditions where the target to be tracked is a projected spot, rather than a spot on *S*'s fingertip. The reason for replication is to develop an apparatus that will enable manipulation of the relevant parameters of both hand and target position and movements. The apparatus is the same one used in all subsequent experiments.

Method

Apparatus.—The *S* was seated in front of an apparatus that allowed simultaneous registration of eye and hand position as well as the projection of an oscilloscope spot that served as the target to be tracked. The *S*'s right arm rested on a flat lever pivoted near the elbow. Movement of the lever was restricted by soft rubber stops to 10° (in visual angle) to the right and left of *S*'s median plane. Registration of arm position was provided by a potentiometer geared to the pivot. The *S* looked down into a box containing a mirror and a ground-glass screen. His head was held in position by having him bite into a bar bearing a wax dental impression. The bite bar was attached to the front of the box, as were the photoresistors used in measuring eye position (described subsequently).

The voltage that was used in monitoring hand position was also fed to the oscilloscope behind the ground-glass screen. The scope trace was projected onto the screen so that the horizontal displacements of the target spot were equivalent to those of the hand. The mirror and screen were arranged so that the virtual image of the spot corresponded closely to the actual position of the tip of the middle finger of *S*'s right hand. Thus, the optical distance of the spot was the same as the actual distance of *S*'s fingertip (when in the median plane) from the cornea of the right eye, 30.5 cm. The screen was kept very dimly illuminated so that the projected scope trace, subtending a visual angle of ⅓°, would be clearly visible. The *S* viewed the target binocularly.

Eye position was monitored using a photoelectric technique. This method, described in detail by Stark, Vossius, and Young (1962), uses a pair of photoresistors (Texas Instruments, LS-400) to measure the amount of diffuse infrared light reflected from the iris-scleral junction. For the experiments reported here, it was accurate to ½° of horizontal displacement of the eye and was approximately linear over a range of 20°. Only the horizontal movements of the right eye were recorded.

Arm (target) position and eye position were recorded simultaneously on two channels of a Sanborn recorder, operating at a paper speed of 100 mm/sec.

Procedure.—For the active condition, *S* was instructed to fixate and follow the moving spot as best he could while rapidly moving his hand back and forth at an irregular rate of approximately 3–4 reversals of direction per second within the 20° travel range of the lever. In the passive condition, *S* relaxed his arm and allowed *E* to move the lever through a pattern of displacements similar to the one *S* used in the active condition. The *S* was again instructed to look at the target at all times. In this and subsequent experiments, approximately 10–15 sec. of recording constituted a trial. The conditions were generally presented in a permuted order and several replications for any condition were usually taken within one session. No order effects were observed in the results of any of the experiments reported here.

Subjects.—Six male and two female employees and students at M.I.T. served as volunteer *S*s. The visual acuity of all *S*s in this and subsequent experiments was sufficient to perform the task without corrective lenses, and, although not formally tested, none had any noticeable degree of strabismus.

Results

Figure 1 shows a portion of one *S*'s tracking records for the active and passive conditions. The hand record is on the top in each pair of records and maximum peak-to-trough travel represents 20°. In the active condition the eye is following the hand quite closely, with a very small latency and few saccades. The contrast with the performance in the passive condition is quite apparent. The passive record is marked by saccades and an increased latency between the eye and hand as compared to that found in the active condition.

Figure 1 is nearly identical to the one presented by Steinbach and Held (1968). There is, in fact, no way of distinguishing records taken under the active condition in the present experiment, where the target was a projected scope trace, from the records

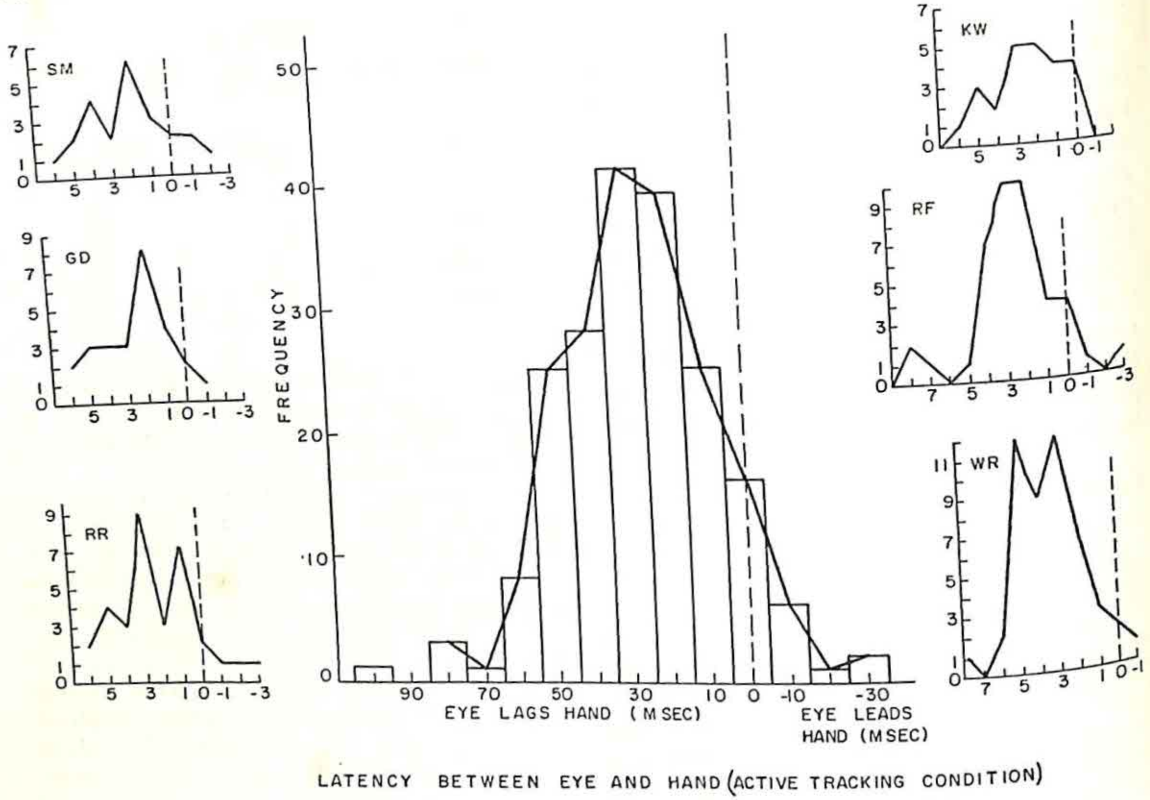


FIG. 2. Frequency histogram of latency between eye and hand for the active tracking condition. (Data for individuals surround the cumulative histogram.)

Subjects and procedures.—Five male employees and students at M.I.T. were *S*s. They were instructed to track as smoothly as possible the target spot which moved with their arms, but which would only appear for a brief (10-msec.) period five times per second. Each *S* was tested under three conditions: active, passive, and "target-only." For the active and passive conditions, the instructions were the same as those for Exp. I. For the target-only condition, *S* removed his arm from the flat lever and had no contact with the apparatus. The *E* moved the lever (and thus the target) in the same fashion under both the passive condition (when *S*'s arm rested on the lever) and in the target-only condition (when *S*'s arm was removed from the lever).

Results

The results for all five *S*s were clear and quite similar to each other. A sample record for one *S* is given in Fig. 3. Looking at the bottom record first, for the target-only condition, we see that minimal tracking occurs. There are some smooth movements of the eye, but they are only weakly correlated with the target and occur only for low rates of movement. When this is compared to the passive record in the middle, a vast improvement can be seen. The tracking is smooth and closely matches the hand movements. There is, however, in the passive record, an increased number of saccades and a slight lag of the eye behind the hand when compared to the active record on top. For the active condition, the tracking was as accurate and smooth as in the case when the target was present continuously and moved by *S*.

For all *S*s the tracking in the target-only condition was a series of saccades with or without some smooth portions, all of which bore little relation to the movement of the target. The perceptual experience was correlated with the tracking performance: *S*s reported seeing a spot flashing on and off in different positions, but having no apparent motion. When the spot was moved by *S* or related to his passively moved limb, the perception was that of a moving target appearing briefly at fixed intervals.

The active/passive differences found in all of the earlier experiments on tracking are still maintained when the target is presented stroboscopically. The tracking in the active condition is more precise and requires few saccades to maintain alignment of eye with target; in the passive condition, there is a lag of the eye behind the hand (and therefore the target) and an increased number of corrective saccades. This is evident from inspection of Table 2.

Discussion

Experiment III demonstrates that the oculomotor system does receive an internal motion signal that will enable it to make smooth eye movements and follow accurately an intermittently appearing target. The tracking performance in the active condition of a target that of

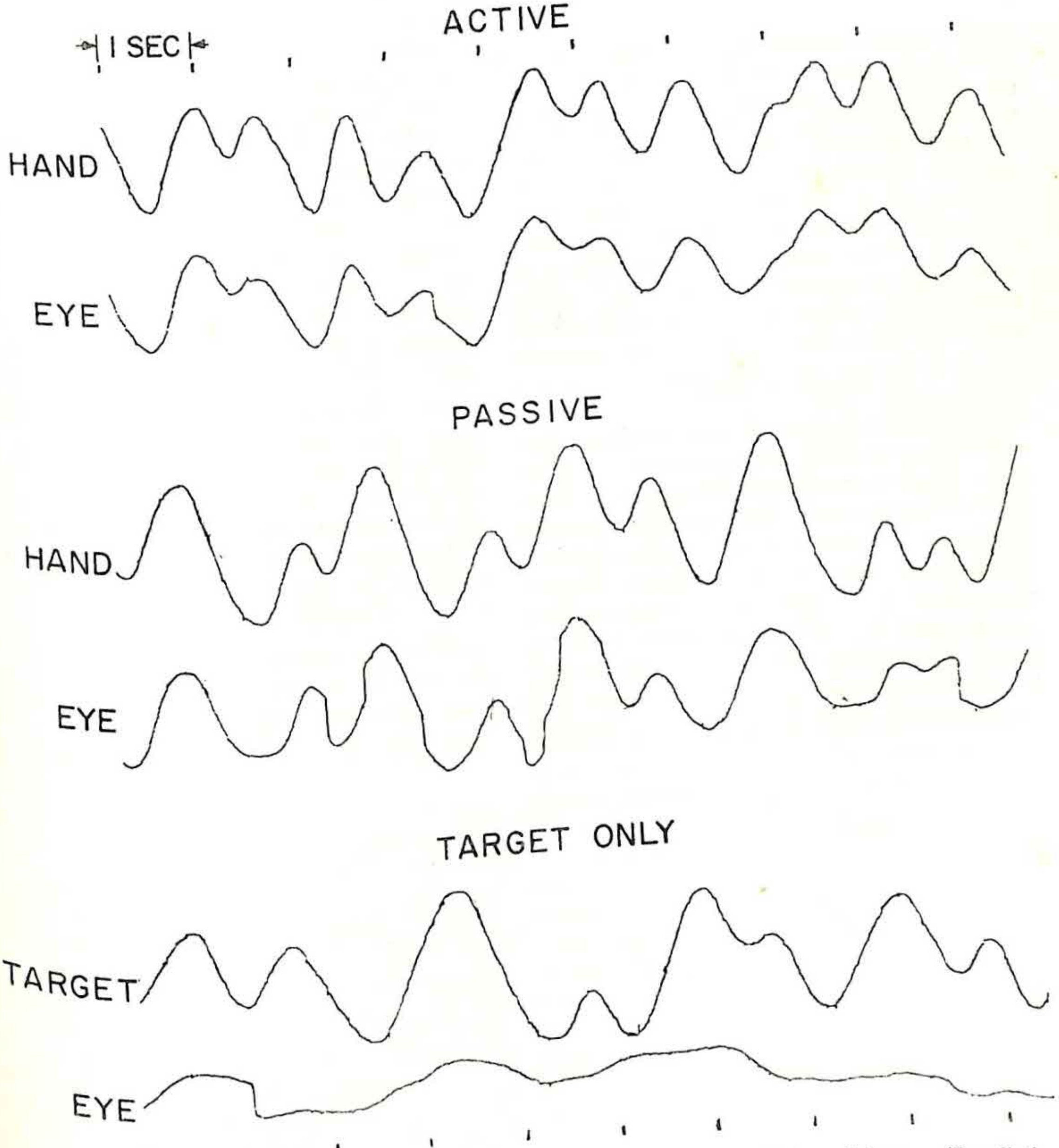


FIG. 3. Eye and hand movement records for active, passive, and "target-only" conditions. (In all three conditions, *S* tracked an intermittently appearing target, on 10 msec. every 200 msec.)

TABLE 2

NUMBERS OF SACCADES OCCURRING IN 10 SEC. OF TRACKING (TARGET PRESENTED INTERMITTENTLY FOR 10 MSEC. EVERY 200 MSEC.)

Ss	Target moved actively	Target moved passively
JS	6	12
JB	25	32
WR	16	31
SM	12	18
MG	8	15
Average	13.4	21.6

itself provides only meagre retinal information about its movements was clear and unequivocal: the performance, in fact, gave no indication that the target was at all degraded. The tracking performance in the passive condition, despite the absence of any outflow, proved to be unexpectedly good (although still poorer when compared to the active condition). Evidently there is enough information from inflow or proprioceptive sources to provide the oculomotor system with a smooth signal to release the pursuit movements.

EXPERIMENT IV

Does the velocity and acceleration information received by the oculomotor system come from the motion of the target or from the efferent signal? This question remains unanswered in the case of a target which is *continuously* present. Experiment IV was undertaken in an attempt to separate these two possible sources by providing *S* with a target whose gross motion was generated by his own arm movements, but which had superimposed on its gross motion a constant sine wave whose frequency was much less than the frequency of the movements of the hand. This type of manipulation provides a target whose turn-around points correspond almost precisely with the turn-around points of the arm (that is, target and arm movements have the same phase), but whose velocity is either greater or less than that of the hand, depending on whether the superimposed sine wave is adding to or subtracting from the instantaneous velocity of the hand. This is made clearer in Fig. 4 showing a sine wave (S) of .5 cycles/sec and approximately one-third the amplitude of the full sweep of the hand motion. The target trace (T) results from the instantaneous addition of the hand (H) and sine voltages. When the hand movement and sine wave are in the same direction, the voltages add and the resulting target trace has a steeper slope at the inflection point than the hand trace, corresponding to a target moving faster than the hand; when the sine and hand voltages are of opposite sign, the slope of the target trace is less than the hand trace, indicating that the hand is moving faster than the target. The addition of the sine, however, has not significantly changed the turn-around points of arm and target.

This situation should allow for a test of whether the eye follows the target or the hand. If the former, then the velocity and acceleration information would seem to be given by the target motion itself, and not by some internally available source. If, however, the eye can be shown to follow the

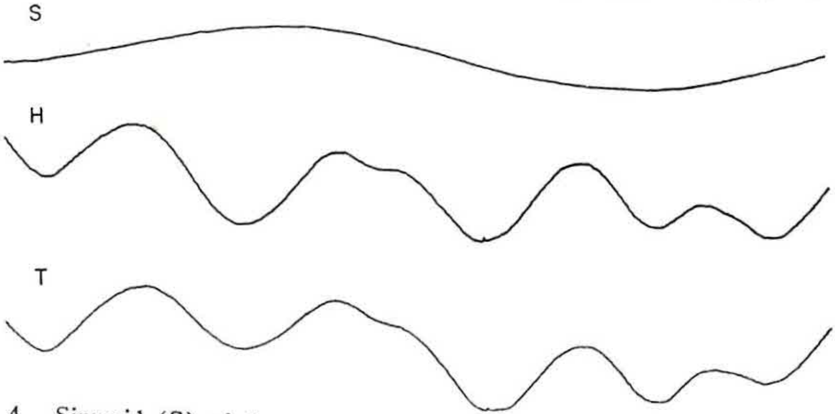


FIG. 4. Sinusoid (S) of .5 cycles/sec superimposed onto hand signal (H) yielding target (T) whose velocity is greater or less than that of the hand.

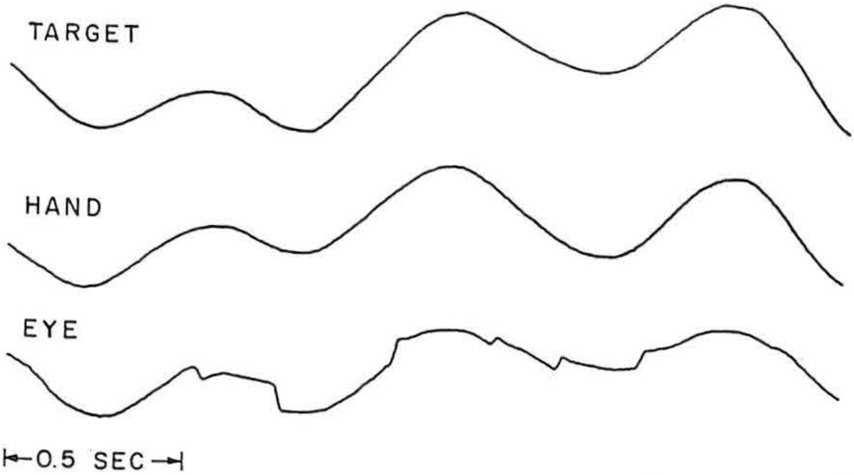


FIG. 5. Portion of a record of target, hand, and eye-movement tracings showing positive saccades (rapid flicks of the eye in the direction of the smooth eye movement) when target velocity is greater than hand velocity, and negative saccades (flicks in the direction opposite to the smooth eye movement) when the hand velocity is greater than the target velocity.

hand or be influenced somehow by its velocity when the hand is voluntarily moved, then one would have strong evidence for the oculomotor system's capacity to use efferent information about velocity and acceleration.

Method

A variable speed, sinusoidally oscillating potentiometer was added to the apparatus. This allowed for the superimposition of a sinusoidal signal on top of the voltage corresponding to hand position on the oscilloscope trace. The potentiometer was adjusted so that the amplitude of the added signal was about one-third the full amplitude of the hand movement.

The experiment was conducted over two sessions. For the first session, four male employees and students were *S*s. They were instructed to follow the target spot as accurately as they could while moving it back and forth irregularly at a rate of 2-4 reversals per second. For all *S*s the oscillating potentiometer was set at .25 cycles/sec. Two of the *S*s were also tested with a superimposed sinusoid of .5 cycles/sec. All the records from this session were only for the active or self-produced movement condition.

In a second recording session, five *S*s (three from the first session as well as one additional male student and one female employee of M.I.T.) were used. In this session, data were collected for passive as well as active conditions of target motions. The *S*s were tested with a superimposed sine of .25 and again at .50 cycles/sec, in permuted order under active and passive conditions.

Results

Figure 5 shows a sample record obtained from one *S* in the active condition where the superimposed signal was .5 cycles/sec. The eye movement record shows several saccades which normally occur when the target motion is too fast for the pursuit system, and increasing positional errors necessitates the use of corrective saccades to reposition the target image on to the fovea. The unusual feature of the records, however, is the presence of saccades which flick the eye in a direction *opposite* to that in which it was moving just prior to and after the saccade. Saccades of this type do occur in normal tracking, but not with the frequency found in these records. Further, these "negative" saccades seem to be occurring whenever the hand is moving faster than the target. The "positive" saccades, which move the eye in the same direction as the smooth movement prior and after the saccade, seem to occur whenever the target is moving faster than the hand.

The frequency and direction of the saccades occurring at or near the inflection point of the eye-movement record were compared with the difference between hand and target velocities. Use of the data from the inflection points insured that a more accurate estimation of the hand and target velocities would be obtained. The cumulative data (a total of 133 points) for the four *S*s of the first session (active condition) are presented in Fig. 6 by the filled circles. Figure 6 confirms the impression given by the in-

GREGORY, R. L. Eye movements and the stability of the visual world. *Nature*, 1958, **182**, 1214–1216.

HELD, R., & FREEDMAN, S. J. Plasticity in human sensorimotor control. *Science*, 1963, **142**, 455–462.

HELD, R., & HEIN, A. Adaptation of disarranged hand-eye co-ordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills*, 1958, **8**, 87–90.

HOLST, E. VON. Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour*, 1954, **2**, 89–94.

HOLST, E. VON, & MITTELSTAEDT, H. Das Reafferenzprinzip. *Die Naturwissenschaften*, 1950, **37**, 464–476.

NOORDEN, G. K. VON, & MACKENSEN, G. Pursuit movements of normal and amblyopic eyes. *American Journal of Ophthalmology*, 1962, **53**, 325–336.

RASHBASS, C. The relationship between saccadic and smooth tracking eye movements. *Journal of Physiology*, 1961, **159**, 326–338.

ROBINSON, D. A. The mechanics of human smooth pursuit eye movement. *Journal of Physiology*, 1965, **180**, 569–591.

SPERRY, R. W. Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, 1950, **43**, 482–489.

STARK, L., VOSSIUS, G., & YOUNG, L. R. Predictive control of eye tracking movements. *Institute of Radio Engineers Transactions on Human Factors in Electronics*, 1962, **HFE-3**, 52–57.

STEINBACH, M. J. Dynamic alignment of eye and hand: The role of efference. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1968.

STEINBACH, M. J., & HELD, R. Eye tracking of observer-generated target movements. *Science*, 1968, **161**, 187–188.

TEUBER, H.-L. The riddle of frontal lobe function in man. In J. M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior*. New York: McGraw-Hill, 1964.

TEUBER, H.-L. Alterations of perception after brain injury. In J. C. Eccles (Ed.), *Brain and conscious experience*. New York: Springer-Verlag, 1966.

(Received March 19, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 377-380

SEARCHING FOR WORDS IN LETTER SETS OF VARYING SIZE[1]

IRA T. KAPLAN,[2] THOMAS CARVELLAS
Ophthalmology Department, New York University Medical Center
AND WILLIAM METLAY
Hofstra University

The *S*'s task was to list all the 4-letter words he could make up from a set of 5-10 different letters. Responses occurred rapidly at first and then more slowly as *S* exhausted his store of words. When *S* had more letters to work with, he could generate a greater total number of words, but took longer to produce any given fraction of them. Thus with 5 letters, *S* found 80% of his limit of 8 words in less than 5 min., whereas with 10 letters, *S* took more than 15 min. to find 80% of his limit of 43 words. The results were analyzed in terms of a model that compared the retrieval of information from memory to a search process.

A situation often arises in which a person must retrieve a particular item of information from among a large number of alternatives. In a visual search task, e.g., *S* must find a target located somewhere on a display containing various symbols or forms. Similarly, in the case of verbal recall, *S* may have to search through many words stored in his memory in order to find a particular target word. Consider an experiment by Bousfield and Sedgewick (1944), who asked their *S*s to list words containing the letters M, T, D. These instructions defined a category of target words and presumably also fixed a larger population of items that *S* searched through. The *S* undoubtedly examined many words that turned out not to be targets: words with some, but not all, of the specified letters. Several other categories, such as "U. S. cities" and "birds," were also investigated, and in every case it was observed that responses were produced most rapidly at first and then more and more slowly as *S* exhausted his supply of target words. This characteristic course of responding suggested that the cumulative distribution of responses could be described by an exponential function of the form

$$F = T(1 - e^{-pt}), \qquad [1]$$

where F is the number of words listed by time t and p is a constant that determines the rate at which the curve rises relative to its asymptote T. The larger p is, the more rapidly the cumulative response curve approaches its limiting value.

The analogy between visual search and the retrieval of information from memory can be formalized in a simple model that compares verbal recall to a search process. Suppose that *S* searches through a set of N items of which T are targets. On every trial, *S* examines one item selected at random. Assume that all items remain in the set after they are examined, so there are always N items in the set. If F is the number of targets that have already been found, then the probability that *S* will find a new target on any single trial is $(T - F)/N$. Thus the rate at which *S* finds targets is equal to $k(T - F)/N$, where k is *S*'s scanning rate in trials or items per unit time. This statement can be written in differential notation as

$$\frac{dF}{dt} = \frac{k}{N}(T - F). \qquad [2]$$

Equation 1 is obtained by substituting

$$p = \frac{k}{N} \qquad [3]$$

and integrating Equation 2. The parameter p, therefore, is the probability per unit time,

---

[1] This study was supported by United States Public Health Service Grant NB-07618 from the National Institute of Neurological Diseases and Blindness.

[2] Requests for reprints should be sent to Ira T. Kaplan, Department of Ophthalmology, New York University School of Medicine, New York University Medical Center, 550 First Avenue, New York, New York 10016.

i.e., the probability density of choosing any particular one of the N items. It is assumed for simplicity that all N items have the same chance of being selected. The effect of variation in probability for different items has been discussed elsewhere (Kaplan & Carvellas, 1969).

The present study used this search model to describe S's behavior in a task that required him to list all the four-letter words he could find in a set of 5–10 different letters. The parameters p and T were estimated by fitting Equation 1 to the cumulative response curves obtained for each number of letters. Since S always searches for four-letter words, his scanning rate k should be constant. Consequently, p should vary inversely with N, the number of items that S searches through. A reasonable hypothesis is that N is equal to the number of four-letter permutations that can be made from the stimulus letters. This quantity can be calculated from the formula

$$N = \frac{L!}{(L-4)!}, \qquad [4]$$

where L is the number of letters that S has to work with. Combining Equations 3 and 4 leads to the prediction that p will decrease as L increases and that each additional letter will produce a progressively smaller decrement in p. The present experiment was designed to test this hypothesis and to investigate the relationship between two measures of S's performance: T, the total number of words he can produce, and p, his relative response rate.

METHOD

Subjects.—The Ss were 48 students of psychology at Hofstra University.

Stimuli.—A set of 5–10 capital letters was hand-printed at the top of 144 sheets of 8 × 12½ in. lined paper. The various sets were based on four different 10-letter sets: (a) RATPELCGYH, (b) LERDOPMNCH, (c) TEDIRMLBFU, and (d) RADILGBNTW. The 5-letter sets consisted of the first 5 letters of each 10-letter set, i.e., RATPE, LERDO, TEDIR, and RADIL. The 6-letter sets comprised the first 6 letters of each 10-letter set, i.e., RATPEL, etc. The 7-, 8-, and 9-letter sets were made up in the same way, with the letters always printed in the same order. Thus there were 24 different sets of letters, and each one was used six times.

Procedure.—The S's task was to list all the 4-letter English words he could make up from the letters at the top of a sheet. He was instructed not to use any letter more than once in the same word and not to list any proper nouns or foreign words. The Ss were tested in two groups of 24, which were matched for their ability to generate 4-letter words from a set of 8 letters. One group received the 5-, 6-, and 7-letter conditions, and the other was given the 8-, 9-, and 10-letter conditions. Each S was tested on three letter sets. Each of these three sets contained a different number of letters, and each was based on a different 10-letter set, so that hardly any solution words were common to more than one of the three sets received by a given S.

The Ss received their three test sheets face down and turned to the first page at a signal from E. In order to mark time on the sheets, Ss drew a line under their last response when E said, "Now." This signal was given at 30 sec., 1 min., and every 1-min. interval thereafter. A total duration of 20 min. was allowed for the 5-, 6-, and 7-letter conditions and 30 min. for the 8-, 9-, and 10-letter conditions. There was a 5-min. rest period between conditions. A 2-min. practice session in which S listed makes of automobiles preceded the experiment proper.

RESULTS

Increasing the number of stimulus letters had two effects on the production of four-letter words: (a) It increased the total number of words that S produced; and (b) it decreased the rate at which S produced any given proportion of the total number of words. The cumulative frequency of words produced as a function of time is shown in Fig. 1. Each point represents the average data for 24 Ss. All the frequency distributions rise rapidly at first and then more slowly as S exhausts his supply of available words. The curves drawn through the distributions are exponential functions of the form of Equation 1. The T values were obtained by visually estimating the asymptotes of the cumulative frequency distributions. The magnitude of p was estimated by drawing a straight line through a rectified plot of log $(T - F)$ against time t. The slope of this line is equal to $-p$.

The distributions are clearly ordered with respect to the total number of words produced. At 20 min. the 5-, 6-, and 7-letter curves are almost horizontal and have nearly reached their estimated T values of 8.0, 14.0, and 20.5 words, where T represents the total number of words S could produce given unlimited time. At 30 min. the 8-, 9-, and 10-letter curves are still rising slowly toward their estimated T values of 28.4, 34.8, and

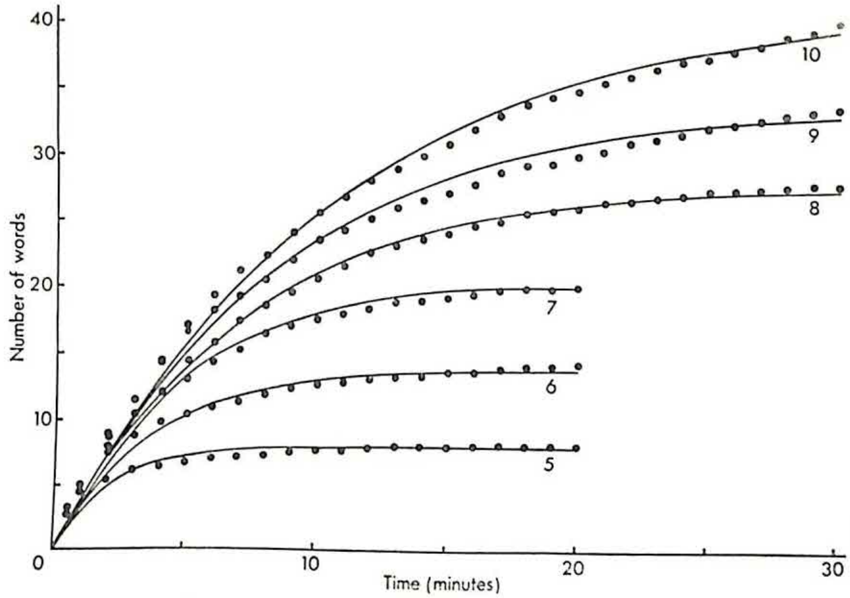


FIG. 1. Cumulative frequency of 4-letter words as a function of time for sets of 5–10 letters. (Each distribution is the average of data from 24 *S*s.)

43.0 words. The increase in T as a function of number of stimulus letters was linear and averaged 7.0 words/letter.

There appears to be a consistent decline in the rate at which each distribution approached its limit. With 5 stimulus letters, the curve rises steeply during the first 5 min. and then levels off. In the 10-letter condition, on the other hand, the curve is still rising steadily even after 30 min. The progressive decline in relative rate of approach is reflected in the decreasing magnitu the probability density p. The p value 5–10 letters, respectively, were .487, .216, .138, .115, and .092 per minute. general, each additional stimulus letter duced a progressively smaller decremen p, as shown by the filled circles in Fig.

DISCUSSION

The probability density p of finding 4- words decreased as the number of stimulus

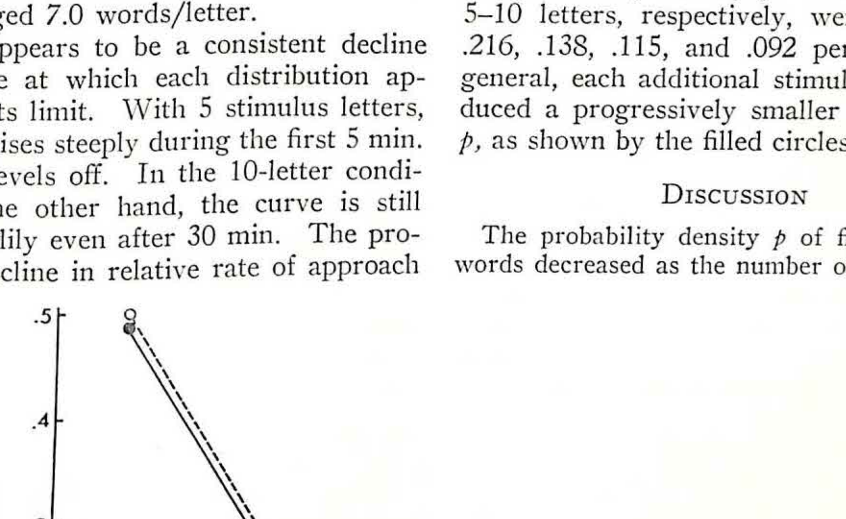


FIG. 2. Probability density p as a function of number of stimulus letters. (The filled circles re sent p values estimated directly from the cumulative distributions of response times. The open cir are values calculated from the formula $p = 4.01/T$.)

ters increased, and the decrement in p produced by each additional letter grew progressively smaller. Qualitatively, this result agrees with the prediction based on the hypothesis that p is inversely proportional to the number of 4-letter permutations that can be made from the stimulus letters. Quantitatively, however, the prediction is inadequate: the observed decrease in p is much less than the hypothesis requires. The magnitude of p for the 5-letter condition (.487) is 5.3 times as large as p for the 10-letter condition (.092). In contrast, the number of 4-letter permutations for the 10-letter condition ($10 \times 9 \times 8 \times 7 = 5040$) is 42 times as large as that for the 5-letter condition ($5 \times 4 \times 3 \times 2 = 120$). Therefore, the hypothesis that S found words by searching randomly through all possible 4-letter permutations of the stimulus letters must be rejected.

In order to modify the original hypothesis, it is proposed that S can exclude many 4-letter permutations from the category of potential words without explicitly examining them. Several studies have shown that in solving anagrams, S does not just rearrange their letters randomly, but instead uses his knowledge of word structure to form those permutations that are most likely to be English words (Kaplan & Carvellas, 1968; Mayzner & Tresselt, 1962; Ronning, 1965). In the present experiment, S might not consider any permutation that consists only of consonants. A more comprehensive but less well-defined rule might exclude all unpronounceable letter sequences. Since the "rules" by which S generates his potential solutions are not known, it cannot be predicted how N varies as a function of the number of stimulus letters. However, an hypothesis that establishes a relationship between p and T can be proposed.

In every condition of the present experiment, regardless of the number of stimulus letters, S's task is the same: to form 4-letter words. When S has more letters to work with, he produces more target words, but he also must examine more nontarget items. Since the same rules generate both target and nontarget items, the ratio of targets T to total number of items N should remain constant. Assume, therefore, that both k, the rate at which S forms and tests items, and T/N, the proportion of items that are words, are invariant. Then substitution in Equation 3 yields $p = C/T$, where C is a constant. If this hypothesis is correct, then the product $pT = C$ should be approximately the same for each number of stimulus letters. Using the results obtained for 5–10 letters, respectively, the product of p and T is found to be 3.90, 3.85, 4.43, 3.92, 4.00, and 3.96, which gives a mean value for C of 4.01. Thus the assertion that p is inversely proportional to T describes the results quite well. This fact is illustrated in Fig. 2, which shows the close agreement between the p values estimated directly from the data and the p values calculated from the formula $p = 4.01/T$, where T was estimated from the data.

When S performs different tasks, k and T/N are likely to vary. There is, however, a connection between T and N that tends to maintain the inverse relation between p and T. The targets are a subset of the set of all items, so that an increase in T produces an increase in N, which in turn causes a decrease in p. Thus even when S recalls words from different categories, a negative correlation between p and T may be expected. Bousfield and Sedgewick (1944) collected data on 14 different verbal recall tasks, e.g., naming birds, automobiles, words containing the letters M, T, D, etc. They reported p and T values (which they labeled m and c) for five categories. The p and T values for the other nine categories were estimated from their latency distributions. The rank order correlation between p and T for all 14 categories was −.64. In the present experiment where S's task was always the same, the rank order correlation between p and T was −1.0. In terms of behavior, this correlation means that when S can recall more words, he takes longer to produce any given fraction of them. In terms of the search model, it means that when there are more targets, there are also more items to examine, so it takes longer to scan any given fraction of the display.

REFERENCES

BOUSFIELD, W. A., & SEDGEWICK, C. H. An analysis of sequences of restricted associative responses. *Journal of General Psychology*, 1944, **30**, 149–165.

KAPLAN, I. T., & CARVELLAS, T. Effect of word length on anagram solution time. *Journal of Verbal Learning and Verbal Behavior*, 1968, 7, 201–206.

KAPLAN, I. T., & CARVELLAS, T. Response probabilities in verbal recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 344–349.

MAYZNER, M. S., & TRESSELT, M. E. Anagram solution times: A function of word transition probabilities. *Journal of Experimental Psychology*, 1962, **63**, 510–513.

RONNING, R. R. Anagram solution times: A function of the "ruleout" factor. *Journal of Experimental Psychology*, 1965, **69**, 35–39.

(Received March 21, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 381–384

IDIODYNAMIC AND POPULAR SETS IN WORD ASSOCIATIONS OF COLLEGE STUDENTS[1]

RONALD NEMAN AND THEODORE R. DIXON[2]

Texas Christian University

A word list was read to the same 176 college students under two sets of instructions: "Give the first word that you think of," followed by "Give the word that most people give." The shift to popular set did not appear related to *S*s' idiodynamic sets which produced predictable influences on set-compatible words under the first instructional set. Additional data regarding idiodynamic sets are presented with regard to continuous associations. Little generality appeared between associations obtained in the single-word and continuous-word association tasks.

The existence of personal word association response tendencies or what Moran (1966) has termed "idiodynamic sets" has been empirically established as a reliable phenomenon. Other investigators have demonstrated that word associations under instructions to give the culturally popular response (populars) tend to move in the direction so instructed, i.e., higher commonality (Horton, Marlowe, & Crowne, 1963; Jenkins, 1959). In an earlier study (Neman, 1968), children were given popular-set word association (PSWA) instructions; those with response sets and who increased commonality scores achieved only 13% of their increases on set-incompatible words. One aim of the present study was to determine if adults could increase commonality on both types of words rather than just those which were set-compatible.

Previously, Moran (1966) observed that association sets facilitated or interfered with responses to respective stimulus words strongly compatible or incompatible with *S*s' sets. Although four different sets have been identified, the described effects have been noted for only the object-referent (OR), concept-referent (CR), and dimension-referent (DR) sets. The present study investigated whether or not such effects would occur when the perceptual-referent (PR) type of words and *S*s were included.

Finally, the present study investigated relations between discrete, free word association (DFWA) response sets and responses obtained as controlled, continuous word associations (CCWA) to different form classes. Moran (1966) observed that in the DFWA experiment, OR, CR, and DR *S*s tended to show preference for noun, verb, and adjective responses, respectively. For the purposes of the present experiment, it was predicted that set-type *S*s would show the same ordering of preference as specified earlier, both in the DFWA and CCWA tasks. The same relation was predicted regarding correlations between semantic and grammatical variables associated with a particular set (e.g., in the DFWA setting, frequency of verbs was expected to correlate most highly with frequency of synonyms and vice versa).

METHOD

Subjects.—The *S*s were 176 students enrolled in introductory psychology classes at the University of Texas at Arlington. They had not served as *S*s in previous psychological experiments.

Word list.—The word list was constructed so that response sets could be inferred by the distribution of associates to the first 48 words and so that the effects of sets could be predicted in *S*s selected as representing sets on the last 36 words. Nine words each were chosen to represent the four idiodynamic sets.

Procedure.—The *S*s were told to write down the first word that they thought of each time they heard *E* call out a word. The 84-word list was then read to *S*s, with a 5-sec. interval between words.

---

[1] The determination of word list frequencies, the scoring of *S* protocols, and the statistical analyses were performed through the facilities of the Texas Christian University Computation Center, A. A. J. Hoffman, Director, Grant U1417-089.

[2] Requests for reprints should be sent to Theodore R. Dixon, Department of Psychology, Texas Christian University, Fort Worth, Texas 76129.

TABLE 1

PREDICTION OF COMMONALITY AND FREQUENCY *z* SCORES

N	Type of *S*	Type of stimulus word/variable							
		PR		OR		CR		DR	
		Frequency[a]	Commonality[b]	Frequency[a]	Commonality[b]	Frequency[a]	Commonality[b]	Frequency[a]	Commonality[b]
31	PR	.40	.33	−.24	−.33	−.15	−.16	−.19	−.30
31	OR	.00	−.36	.11	.12	.06	.11	−.07	−.12
18	CR	.05	.09	−.30	.27	.24	.10	.06	.01
35	DR	−.52	−.47	.21	−.08	−.04	−.01	.48	.38

Note.—Abbreviated: PR = perceptual-referent; OR = object-referent; CR = concept-referent; and DR = dimension-referent.

[a] Frequency scores *on variables* representing sets in the second part of the list.

[b] Commonality scores *on words* representing sets in the second part of the list.

After the word list was completed, *S*s were told that some of the same words would be called out again and that this time they should write down the word that most people would think of when they heard the word. The words in the second part of the list were then called out again.

Finally, *S*s were told that the name of a part of speech would be written at the top of the next page of their booklets. The part of speech would be either a noun, verb, or adjective. The *S*s were to write down as many examples as they could of the part of speech indicated. After 1 min., *S*s were stopped, asked to turn the page, given the preceding instructions, and allowed to write for another minute. After the second minute, the previous procedure was repeated. On each trial, *S*s had a different part of speech. Materials were arranged so that all six possible orders would be sampled.

Variables.—Frequency scores were obtained by summing the number of times each of the following categories were assigned to each *S*: (*a*) functions, (*b*) predications (PR associates), (*c*) related objects (OR associates), (*d*) synonyms, (*e*) superordinates (CR associates), (*f*) contrasts, and (*g*) coordinates (DR associates). Definitions for these categories are provided in Moran and Murakawa (1968).

Commonality was the frequency with which the sample gave each response word. Part of speech categories, noun, verb, and adjective, were assigned to the stimulus and response word.

Scoring.—Frequency tabulations of all responses were compiled for responses with commonality equal to two or more; scoring was then assigned to these responses by reference to prescored norms for the same words. These norms were collected and scored by Moran (1967). When the part of speech was indeterminable from the context of the stimulus and response word, no scoring was entered. Similarly, for logical or semantic variables, when none of the categories or more than one category fitted the association, no scoring was done.

Statistical analysis.—All correlation coefficients reported are Pearson product-moment coefficients. All correlations and *z* scores were based on means and standard deviations obtained in the sample ($N = 176$).

RESULTS AND DISCUSSION

Effects of set in DFWA.—The proportion scores of each *S* were converted into *z* scores, and then set-representative *S*s were selected on the basis of the following criteria: Set-type *S*s were those having *z* scores greater than .5 on variables representing one set and less than .5 on variables representing other sets. It was predicted that *S*s would have their highest commonality *z* scores on those words among the last 36 most compatible with their sets. With the same group of *S*s, it was predicted that those high on variables representing one set on the first part of the list would also have their highest frequency (proportion) *z* scores on variables representing the same sets in the second part of the list. In Table 1 it can be seen that PR as well as other set-type *S*s achieved their highest commonality and frequency *z* scores on words and variables compatible with their sets. The exact probability of predicting the highest value in the diagonal was .004.

Effects of set in PSWA.—It should be mentioned that previous findings (Horton et al., 1963; Jenkins, 1959) with adults were confirmed regarding increased commonality due to instructional effects. Mean commonality scores increased from 2,689.4 under DFWA to 3,138.3 under PSWA instruction,

$t(175) = 8.1$, $p < .001$. Analysis of increases in commonality scores by type of *S* and word revealed that, overall, 78.5% of the total increase was due to increases on nonset-type words. Increases appeared in all sets for all types of words, whereas with children, the mean commonality scores actually declined on some incompatible words. In adults, it was more often the case that greater increases occurred on words compatible with some set other than that of the *S*. Considering this latter fact, and considering that there was a maximum or ceiling value for each of the four classes of stimulus words, it was possible that on set-compatible words *S*s were so near this maximum that upward changes on these words would be restricted and negligible. However, the mean increase observed on set-compatible words equalled 9.2% of the maximum and 10.8% of the maximum for incompatible words, the mean difference not being significant. Nevertheless, even granting the operation of a ceiling effect would not account for the increases on set-incompatible words, which was the major way this sample differed from children; i.e., both groups increased commonality on set-compatible words. These findings provide support for the contention that two different association hierarchies are manifested during the DFWA experiment, one representing the personal set of *S* and the second representing a culturally determined set (Moran & Murakawa, 1968; Swartz & Moran, 1968).

Correlations between grammatical and semantic variables.—As predicted, nouns and related objects correlated highest with each other ($r = .80$), as did synonyms and verbs ($r = .40$), and contrast and adjectives ($r = .47$). These last two values jumped to .59 and .75 for synonyms and contrasts, respectively, when correlated with paradigmatic verb and adjective responses.

Effect of sets on grammatical variables in DFWA and CCWA.—In Table 2 are shown mean *z* scores for set-type *S*s on frequency of nouns, verbs, and adjectives given in DFWA and CCWA. Regarding the DFWA task, while OR and CR *S*s had their highest means on nouns and verbs, respectively, DR *S*s were highest not on adjectives but on verbs. It is possible that DR *S*s failed to achieve their highest score on adjectives because of a paradigmatic restriction against giving adjectives to anything but adjectives. It is also possible that proportion scores would better reflect the differential reliance on form class across sets. In addition, because the frequency count of responses by form class included all words in the list, because some of the same words were used to select *S*s, and because of the correlation between form class and semantic variables, possible biases were introduced which would favor the predictions made. Thus the results regarding the effect of set on form class in DFWA must remain tentative.

More inconsistent with expectation were the results for the CCWA task; with regard to each form class, both PR and DR *S*s were above the mean (had positive *z* scores), and the OR and CR *S*s were below the mean (had negative *z* scores).

These results were considered surprising for the following reasons: The PR and DR

TABLE 2

FREQUENCY SCORES BY FORM CLASS IN DFWA AND CCWA TASKS

N	Type of *S*	Form class					
		Nouns		Verbs		Adjective	
		DFWA	CCWA	DFWA	CCWA	DFWA	CCWA
31	PR	−.52	.35	.10	.37	.03	.08
31	OR	.75	−.20	−.06	−.02	−.57	−.25
18	CR	−.25	−.06	.23	−.41	.22	−.09
35	DR	−.19	.30	.27	.01	.11	.29

Note.—Abbreviated: PR = perceptual-referent; OR = object-referent; CR = concept-referent; and DR = dimension-referent.

*S*s have consistently appeared at opposite ends of the "associative spectrum" in previous studies; the predication and contrast-coordinate variables are negatively correlated consistently; PR *S*s tend to give syntagmatic, low-commonality responses with high latencies, while DR *S*s tend to do just the reverse. Factor analytic studies of semantic variables have shown that PR and DR variables load at opposite ends of one factor, and when the number of factors extracted is two, the OR and CR variables load on opposing poles of this second factor.[3] It appears that what might be called divergent response styles in the free word situation become convergent styles in the continuous association task.

Inasmuch as DFWA and CCWA responding do not appear to follow the same rules, theoretical accounts need to consider separately each type of associative behavior. If the production of continuous associations represented but a serialization of the production of single associations, it could be expected that measures of "associative production" in the two cases would be correlated. But when commonality scores obtained in DFWA were correlated with frequency scores obtained in CCWA, the correlation proved low and not significant ($r = .06$).

The present experiment was generally supportive of previous findings on associative sets: (*a*) findings can be generalized across all sets regarding the facilitating and interfering effects of sets; (*b*) support for the two hierarchy association hypotheses was gained; and (*c*) grammatical variables were shown to correlate as predicted with logical association categories. Extension to other domains was only partly successful, and additional studies are necessary regarding the effect of set on grammatical variables in both the DFWA and CCWA tasks. Also different methods for estimating production measures would be useful in comparing associations across types of associative tasks. Such research holds promise for contributing to an account of cognitive factors operating in associative behavior.

REFERENCES

Horton, D. L., Marlowe, D., & Crowne, D. P. The effect of instructional set and need for social approval on commonality of word association responses. *Journal of Abnormal Social Psychology*, 1963, **66**(1), 67–73.

Jenkins, J. J. Effects on word-association of the set to give popular responses. *Psychological Reports*, 1959, **5**, 94.

Moran, L. J. Generality of word-association response sets. *Psychological Monographs*, 1966, **80**(4, Whole No. 612).

Moran, L. J. Objective approximations of idiodynamic sets. *Psychological Reports*, 1967, **20**, 827–833.

Moran, L. J., & Murakawa, N. Japanese and American association structures. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 176–181.

Neman, R. Instructions, commonality, and word association sets. *Psychological Reports*, 1968, **22**, 1078.

Swartz, J. D., & Moran, L. J. Association structures of bright children at ages nine and twelve. *Multivariate Behavioral Research*, 1968, **3**, 189–198.

(Received March 21, 1969)

---

[3] Extraction of a third factor usually separates the OR and CR variables.

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 385–389

EFFECTS OF AROUSAL ON HIERARCHICALLY ORGANIZED RESPONSES[1]

A. STEVEN FRANKEL[2]

University of Southern California

One hundred fifty males served as *S*s in an investigation of the Broen and Storms reaction potential ceiling hypothesis. Two- and three-alternative hierarchies of responses were "built in" with a paired-associate learning procedure, and as dynamometer-induced arousal increased, the frequency of dominant responses decreased, with corresponding increases in the frequency of competing responses.

Broen and Storms (1961) discussed schizophrenic deficit in Hullian terms and postulated a "reaction potential ceiling." According to this view, there is a maximum level of reaction potential (*E*), which is joint function of drive (*D*) and habit strength (*H*). The *E* ceiling lies at a level below the product of *D* maximum and *H* maximum.

Given any stimulus or stimulus situation and a set of associated responses, an increase in *D* would increase *E* for all responses, with the strongest or dominant response increasing at the most rapid rate. If *D* is strong enough to bring the dominant response to *E* ceiling, any further increase in *D* would serve to increase the probability of occurrence of competing responses, while decreasing the probability of occurrence of the dominant response. This formulation thus accounts for the inverted-U function found between arousal and performance (e.g., Malmo, 1959) and has been applied theoretically to such psychopathological behaviors as association disturbance, concept disturbance, delusions, hallucinations, and regression (Broen & Storms, 1966).

Empirically, increased arousal (accomplished via induced muscular tension) has been shown to decrease responding of normal males to a training stimulus, while increasing generalization (Broen, Storms, & Goldberg, 1963). Broen and Storms (1964), using psychiatric patients, also reported increases in arousal resulting in a decreased number of appropriate responses and a concomitant increase in competing responses in a discrimination task. Finally, arousal has been shown to be related to paired-associate learning (Spence, 1956) and the eyeblink response (Lovaas, 1960).

The *E* ceiling formulation of how nondominant responses may become increasingly probable as a function of increasing arousal level leads to some specific predictions regarding the effects of increasing arousal on paired associates in which each stimulus has more than one associate. The present research dealt with the effects of arousal on hierarchically organized responses. Nonsense syllables were used as stimuli and responses in order to insure that the probabilities of the responses would be under experimental control. Response hierarchies were "built in" through the use of a paired-associate learning technique in which each stimulus was paired with more than one response. The frequencies of presentation of each response with its stimulus were varied so as to result in a heirarchical organization (Brown, Osterhout, & Voss, 1962; Goss & Sugerman, 1961; Voss, Thompson, & Keegan, 1959).

The response hierarchies contained either two or three responses, and arousal was manipulated via muscular tension (e.g., Broen & Storms, 1964; Malmo, 1959) and

---

[1] This study was submitted to the graduate school of Indiana University in partial fulfillment of the requirements for the PhD degree. The author is deeply indebted to A. M. Buchwald, who chaired his thesis committee and provided invaluable aid in the preparation of the manuscript.

[2] Requests for reprints should be sent to A. Steven Frankel, Department of Psychology, University of Southern California, Psychological Research and Service Center, 734 West Adams Blvd., Los Angeles, California 90007.

TABLE 1

TRAINING MATERIAL: FREQUENCIES OF APPEARANCES OF RESPONSES WITH STIMULI

Approximate training proportion	Two-alternatives responses		Three-alternatives responses		
	A_1 (dominant)	A_2 (competing)	B_1 (dominant)	B_2 (competing)	B_3 (competing)
.9-.1	45	5	45	5	5
.7-.3	35	15	35	15	5
.5-.5	25	25	25	25	5

electric shock[3] (Kuethe & Eriksen, 1957; Schnore, 1959), to test the hypotheses that (*a*) increasing arousal would result in decreasing frequencies of dominant responses, with concomitant increasing frequencies of competing responses, and (*b*) the effects of arousal would be greatest when the habit strengths (training frequencies) of the dominant responses were greatest.

METHOD

Subjects.—One hundred and fifty males with no history of psychiatric involvement were drawn from introductory courses in psychology at Indiana University to serve as *S*s as part of a course requirement.

Apparatus.—A hand dynamometer was constructed from a cable tensiometer and was wired to a servographic recorder such that the strength of pull exerted by *S*s could be recorded in terms of kilograms. A bell was wired to the dynamometer so that it would ring if the pressure indicator varied above or below a given value by 1 kg. (after Broen et al., 1963). The dynamometer was used in connection with the arousal variable.

Material.—Twenty-one CVC trigrams, chosen on a random basis from the middle range of meaningfulness on Archer's (1960) list, served as stimulus material. The nonsense syllables were typed on 3 × 5 in. cards.

Training decks.—Six nonsense syllables served as stimuli (see Table 1). Three of these (A stimuli) were used with two-alternative hierarchies, and three (B stimuli) were used with three-alternative hierarchies.

Nine training decks were constructed. Each deck was comprised of 105 cards, 50 of which showed an A stimulus with either an A_1 or an A_2 response and 55 of which showed a B stimulus with either a B_1, B_2, or B_3 response.

Three different proportional relationships between dominant (A_1; B_1) and competing (A_2; B_2, B_3) responses were used. Three decks were constructed in each of the three response proportion conditions (.9–.1, .7–.3, and .5–.5) (see Table 1). Order of presentation of cards in all decks was randomly determined.

Testing decks.—Associated with each training deck was a testing deck, which consisted of 30 cards. On each card appeared one of the two stimuli used in the training deck and all the responses which had been presented with that stimulus. Fifteen cards showed the two-alternative stimulus with its two responses (A_1 and A_2), and 15 cards showed the three-alternative stimulus with its three responses (B_1, B_2, and B_3).

Procedure.—The *S*s were randomly divided into 15 groups of 10 (arousal by response proportions). The mean maximum dynamometer pressure over a 30-sec. period was determined for each *S* prior to training.

Training.—All *S*s were then presented with three training decks in counterbalanced order and were instructed to pronounce each stimulus–response pair to control for attention.

Testing.—A testing deck was presented to *S* in between each training deck. The *S*s were instructed to "choose" one of the responses which appeared with the stimulus on each card in the testing deck. The *S*s were given no rationale for making their choices, and all questions were answered as follows: "Just choose one of the responses on each card. Do the best you can." No feedback was given to *S*s regarding the "correctness" of their choices.

During testing, three levels of arousal were employed: high, medium, and low. Just prior to the presentation of each testing deck, *S*s in the high-arousal condition were instructed to grasp the dynamometer and squeeze it with half the average pretest pressure. The *S*s in the medium-arousal condition were told to use one-fourth of the pretest pressure, and *S*s in the low-arousal condition were told to place their hand on the device, but not to apply any pressure. For the high- and medium-arousal condition, a bell sounded if the pressure varied above or below the individually determined value by 1 kg.

In order for the effects of arousal to dissipate between testing and the subsequent training on new stimulus material, all *S*s were given a deck of playing cards after each testing deck, were instructed to shuffle the deck, and then to turn the cards face up, one at a time, while calling out "ace, two, three," etc. Each time a turned-up card corresponded with the number uttered by *S* he had to reshuffle and start again. This filler task lasted for 4 min., which, according to Bourne (1955), should have been enough time for the effects of arousal to dissipate.

---

[3] Two shock groups (high and low) were also run in this study. The resulting data were very similar to those resulting from dynamometer-induced arousal and they have been omitted from the present study in order to achieve a briefer and clearer presentation.

RESULTS

For purposes of convenience, the results are arbitrarily presented in two parts: dominant response data and competing response data. Nevertheless, it should be noted that the total frequency of competing responses in the hierarchies in the present study is perfectly negatively correlated to the frequency of appearance of dominant responses.

Dominant responses.—The proportions of dominant responses produced during testing by *S*s in both two- and three-alterantive conditions are shown in Table 2. These data were treated with an analysis of variance.

The proportion of dominant responses on test trials was positively related to the frequency with which they were shown during training, $F = 29.72$, $p < .01$, and was greater for two alternatives than for three alternatives, $F = 22.99$, $p < .01$. Increased amounts of dynamometer pull (arousal) during testing were associated with decreasing proportions of dominant responses in testing, $F = 72.00$, $p < .01$, with the decremental effect of increased arousal being positively related to the frequency of dominant responses during training, $F = 3.61$, $p < .01$.

Under low dynamometer pull, the proportions of dominant responses made in testing closely matched the proportions with which these were presented during training. (I dividual *t* tests comparing proportions dominant responses during training an testing under low arousal uniformly faile to show any significant differences.) Unde increasing levels of dynamometer pull, thi matching became increasingly distorted an the proportions tended to converge.

TABLE 2

MEAN PROPORTION OF DOMINANT RESPONSES UNDER THREE LEVELS OF DYNAMOMETER-PRODUCED AROUSAL

Training proportion	Level			
	Low	Medium	High	*M*
Two alternatives				
.90 (45)	.869	.701	.654	.743
.70 (35)	.687	.622	.584	.631
.50 (25)	.521	.514	.545	.527
M	.6923	.6123	.5943	
Three alternatives				
.82 (45)	.847	.595	.529	.657
.64 (35)	.662	.539	.538	.580
.46 (25)	.479	.487	.437	.468
M	.6627	.5403	.5013	

Note.—Training frequencies are given in parentheses.

TABLE 3

MEAN PROPORTION OF COMPETING RESPONSE UNDER THREE LEVELS OF DYNAMOMETER-PRODUCED AROUSAL

Training proportion	Level			
	Low	Medium	High	*M*
Rseponse B_2				
.09 (5)	.085	.242	.258	.191
.27 (15)	.250	.282	.322	.284
.46 (25)	.441	.425	.383	.416
M	.2587	.3163	.3210	
Response B_3				
.09 (5)	.068	.173	.213	.1513
.09 (5)	.088	.179	.150	.139
.09 (5)	.080	.088	.180	.116
M	.0787	.1467	.1810	

Note.—Training frequencies are given in parentheses.

Competing responses.—The proportions o competing responses produced in testing i the three-alternative hierarchies are show in Table 3. An analysis of variance of thes data (amount of dynamometer pull by number of competing responses during training by response alternative, B_2 or B_3) was performed. In this analysis, the .5–.5 condition was dropped, since one of the "competing responses" (B_2) was actually identical in terms of frequency of presentation during training to the "dominant" response.

As may be seen in Table 3, the proportion of competing responses in testing increased with increasing arousal, $F = 11.71$, $p < .01$. Although the results of the analysis of variance indicated that B_2 was given significantly more often than B_3 overall, a *t* test on B_2 and B_3 in the .9–.1 condition yielded no significant differences. Further, a significant interaction between number of competing responses in training and response alternative (B_2 or B_3), $F = 44.33$, $p < .01$, suggests that the difference between B_2 and B_3 is greater in the .7–.3 condition than in the .9–.1 condition. Finally, the effect of dynamometer pull in the proportion of competing responses depended jointly on the frequencies of presentation during training and the response alternative, $F = 4.124$, $p < .05$. The differences between B_2 and B_3 at different levels of dynamometer pull vary more in the .7–.3 condition than in the .9–.1 condition.

DISCUSSION

A major implication of the Broen and Storms (1961) concept of a "reaction potential ceiling" is that with increased levels of arousal, assuming that habit strengths of dominant responses are sufficiently high, there should be less frequent usage of dominant responses. It will be recalled that this expectation is based on the view that there is a maximum level *E* which is a joint function of drive (arousal) and habit strength.

As may be seen in Table 2, arousal has had the general effect of decreasing the frequency of dominant responses in both two- and three-alternative hierarchies of responses.

A second major implication of the *E* ceiling hypothesis is that the decremental effect of arousal should be greatest when the habit strength of the dominant response is greatest. Thus, it might be expected that the responses which appeared 45 times in training would reach ceiling more rapidly than those which appeared 35 times, which in turn would reach ceiling more rapidly than those which appeared 25 times. The general effect would be a smaller decrease for those dominant responses which appeared 35 times in training than for those which appeared 45 times, with those which appeared 25 times showing the smallest decrease. As may be seen in Table 2, the decrease in frequency of dominant responses in testing was greater at the highest training level than at the medium or lowest training levels.

In general, then, it may be said that the predictions made with regard to the effects of arousal on dominant and competing responses have been supported. As arousal increased, the frequency of dominant responses decreased, with a corresponding increase in the frequency of competing responses. Further, the effects of arousal were greatest when the habit strength of the dominant response was greatest.

In most of the literature dealing with the effects of arousal on performance, the "inverted-U" function is mentioned (e.g., Broen & Storms, 1961; Malmo, 1959). What is suggested by the "inverted-U" function is that as arousal increases, the frequency of "dominant," "correct," or otherwise appropriate responses increases to a maximum. As arousal continues to increase beyond this point, the frequency of such responses decreases. Such an occurrence is certainly expected for dominant responses which are low in *H*, since they presumably would have to increase further in *E* to reach ceiling than do those responses which are already relatively high in *H*.

In the present study, dominant responses were less frequently emitted in testing as arousal increased, regardless of training frequency. This finding might be interpreted to suggest that "medium" arousal, induced by squeezing a dynamometer with one-fourth of one's force, is enough to bring responses to ceiling which, in terms of frequency in training, are relatively low in *H*. This seems unlikely, but the ceiling hypothesis might be strengthened if it were shown that there is a level of arousal which would serve to increase the frequency of dominant responses. Perhaps a level of arousal less than one-fourth maximum force is necessary, but this is an empirical question.

Finally, the results of the present study seem to agree with those of past studies in the area of paired-associate learning with varying percentages of alternative responses. Further, since previous studies (e.g., Voss, et al., 1959) have used the anticipation method, while the present study used the method of simultaneous presentation, it may be said that these results occur independently of the method employed.

REFERENCES

ARCHER, E. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74**(10, Whole No. 497).

BOURNE, L. E. An evaluation of the effect of induced tension on performance. *Journal of Experimental Psychology*, 1955, **49**, 418–422.

BROEN, W. E., & STORMS, L. H. A reaction potential ceiling and response decrements in complex situations. *Psychological Review*, 1961, **68**, 405–415.

BROEN, W. E., & STORMS, L. H. The differential effect of induced muscular tension (drive) on discrimination in schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1964, **68**, 349–353.

BROEN, W. E., & STORMS, L. H. Lawful disorganization: The process underlying a schizophrenic syndrome. *Psychological Review*, 1966, **73**, 265–279.

Broen, W. E., Storms, L. H., & Goldberg, D. H. Decreased discrimination as a function of increased drive. *Journal of Abnormal and Social Psychology*, 1963, **67**, 266–273.

Brown, A., Osterhout, D., & Voss, J. F. The effect of one probabilistic verbal association upon the concomitant acquisition of a second association. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 48–53.

Courts, F. A. Relations between muscular tension and performance. *Psychological Bulletin*, 1942, **39**, 347–367.

Goss, A. E., & Sugerman, M. F. Paired associates learning with varying percentages of occurrence of alternative response members. *Journal of Experimental Psychology*, 1961, **62**, 24–34.

Kuethe, J. F., & Ericksen, C. W. Personality, anxiety, and muscle tension as determinants of response stereotypy. *Journal of Abnormal and Social Psychology*, 1957, **54**, 400–404.

Lovaas, O. I. The relationship of induced muscular tension, tension level, and manifest anxiety in learning. *Journal of Experimental Psychology*, 1960, **59**, 145–152.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychological Review*, 1959, **66**, 367–387.

Schnore, M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *Journal of Experimental Psychology*, 1959, **58**, 117–128.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale University Press, 1956.

Voss, J. F., Thompson, C. P., & Keegan, J. H. Acquisition of probabilistic paired associates as a function of S-R_1, S-R_2 probability. *Journal of Experimental Psychology*, 1959, **58**, 390–399.

(Received March 26, 1969)

ERRATUM

In the article, "Information Processing as a Function of Speed versus Accuracy," by James M. Swanson and George E. Briggs (*Journal of Experimental Psychology*, 1969, **81**, No. 2, 223–229), the reaction time scale for the $H_c = 1.0$ data of Fig. 2 should range from .300 to .700 sec. and not from .200 to .600 sec. as reported. The regression equation $RT = .353 + .154\ (H_t)$ is correct.

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 390–392

SUPPLEMENTARY REPORTS

LONG-TERM RETENTION OF MODALITY- AND NONMODALITY-SPECIFIC HABITUATION OF THE GSR[1]

GORDON B. HARDING[2] AND GARY R. RUNDLE

New Mexico State University

Twenty-four human *S*s received 20 stimulus presentations of either the same modality (Same *S*s) or a different modality (Changed *S*s) on each of 3 successive wk. Overall GSR magnitude decreased from the first to the second week for both groups, but a reliable decrement on Trial 1 with weeks appeared only for the Same *S*s. When skin resistance level (SRL) was used as a measure of arousal, Changed *S*s became less aroused with amount of prior modality change.

The present research compared the long-term retention of modality-specific and nonmodality-specific habituation of the GSR without foreknowledge of the stimuli to be employed. One group of *S*s received 20 trials of the same stimulus on the same day of each of 3 successive wk., while a second group received a stimulus of a different modality on each of those same days.

The skin resistance level (SRL) at the time of stimulus onset was also obtained to determine the changes in level of arousal (Malmo, 1962) that accompany habituation to old and new stimuli.

Method.—The *S*s were 48 introductory psychology students, who volunteered for the experiment to satisfy partially a course requirement. A light, tone, and vibrator served as stimuli. The light was white and came from a 7 × 4 in. source 4 ft. from *S* at approximately eye level. The tone was a 1,000-cps signal at 78 db. (.0002 dynes/cm^2) and was produced by a Hewlett-Packard 200 AB audio oscillator and fed through Willson sound barrier earmuffs. The vibrator was made by Vogt Appliance Company, Kalamazoo, Michigan, and was attached to *S*'s right leg by rayon straps over a $\frac{1}{16}$ in. layer of foam rubber. The two vibrator heads made contact with the calf of the leg.

The light, tone, and vibrator were judged to be of equivalent intensity prior to the experiment. The straps and foam rubber arrangement and the earmuffs reduced the audible noise of the vibrator to a faint hum. To that extent, the modalities of the three stimuli were nonoverlapping.

The GSR was measured by a Fels dermohmmeter Model 22A, with $\frac{3}{4}$-in. diameter zinc sulfate, agar powder, and distilled water.

The 48 *S*s were assigned randomly to either the changed-modality or the same-modality group. Of the 24 *S*s in the Changed group, 4 were assigned to each of the six possible orders of the three different stimuli, and *S*s were given 20 trials of a different stimulus on the same day of the week for 3 successive wk. The 24 *S*s in the Same group were assigned randomly to the three stimuli, and each *S* received the same stimulus, either the light, the tone, or the vibration, on the same day of the week for 3 successive wk.

On entering the laboratory, *S*s were seated and the GSR electrodes were attached to the dorsal and palmar surfaces of the left hand. The *S*s were told that some stimuli were to be presented, but they were never informed by instruction which stimulus was to occur on any given day, nor were they even informed that a tone, light, or vibration could occur.

The vibrator and earphones were attached to each *S* each week, regardless of the condition scheduled for that *S*.

Following the instructions, *E* switched off the lights and closed the door to the *S* room. Twenty presentations of the same stimulus were then given at a mean intertrial interval of 40 sec. with a range of 20–60 sec. Stimulus duration was 1 sec. Each *S* was run on the same day and at the same time by the same *E* for 3 consecutive wk., except three *S*s who were run by a different *E* on Week 3.

The GSR was read to the nearest 200 ohms and was defined as $\log_{10} [(1/R_a - 1/R_b)\, 10^8 + 1]$, where R_b = the skin resistance in ohms at the onset of the stimulus and R_a = the minimum skin resistance in ohms within approximately 4 sec. of stimulus onset.

Results and discussion.—Figure 1a shows the mean GSR magnitude on each of the 20 trials for Weeks 1, 2, and 3 for the Same group. Analysis of variance of the GSR magnitude/*S*/trial/wk showed the effects of trials, F (19, 437 = 13.35, weeks, F (2, 46) = 12.92, and Trials × Weeks, F (38, 874) = 2.26, to be reliable with a $p < .01$.

Evidence for long-term retention of habituation was obtained by analysis of the weeks and Trials × Weeks effects at the simple level (Tukey *a* procedure, Winer, 1962). There was a reliable decrement, $p < .01$, in the mean GSR magnitude from (*a*) Week 1 to Week 2, but not thereafter, and (*b*) Trial 1 on Week 1 to Trial 1 on Week 2.

The reliable decrement in first-trial response magnitude from Week 1 to Week 2 suggests that

---

[1] This research was supported by the Claude Dove Research Fund of New Mexico State University.

[2] Requests for reprints should be sent to Gordon B. Harding, Department of Psychology, New Mexico State University, Box 3452, Las Cruces, New Mexico 88001.

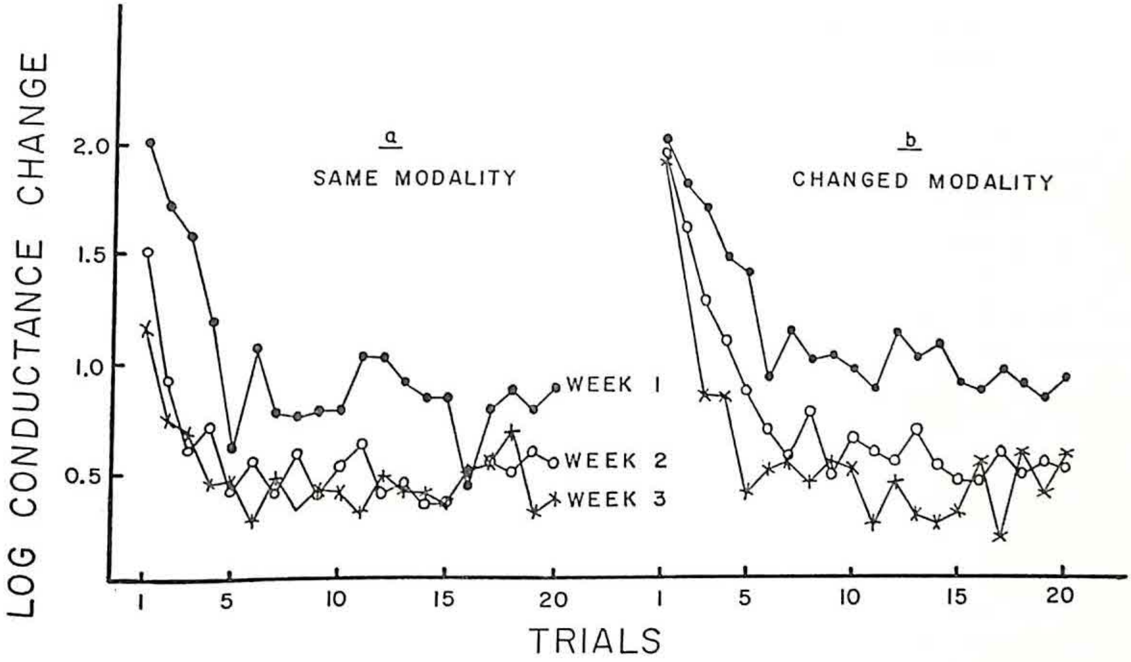


FIG. 1. Mean GSR magnitude/trial/wk for the same-modality and changed-modality Ss.

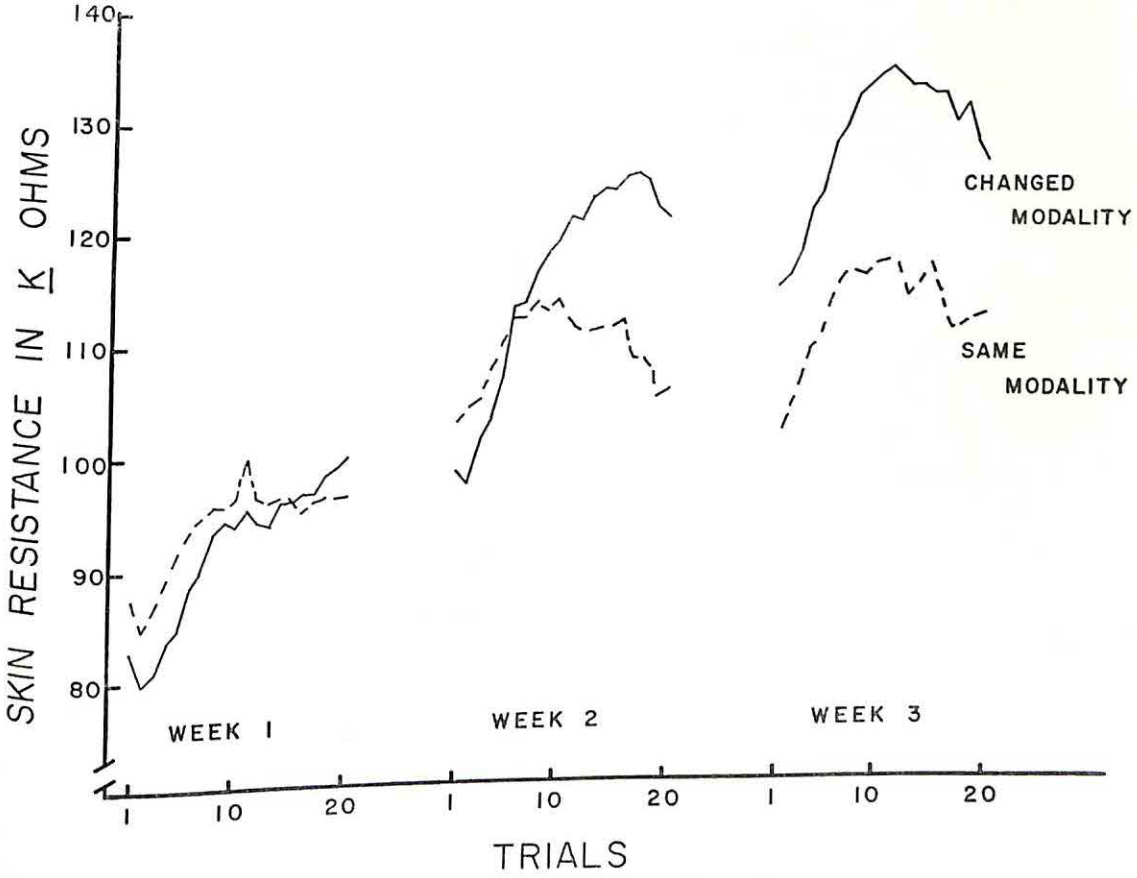


FIG. 2. Mean SRL/trial/wk for the Same and Changed groups.

proprioception is not a necessary condition for retention of habituation (Kimmel & Goldstein, 1967).

Figure 1b shows the mean GSR magnitude on each of the 20 trials for Weeks 1, 2, and 3 for the Changed group. Analysis of variance of the GSR magnitude/S/trial/wk showed the effects of trials, $F\ (19, 437) = 30.88$, and weeks, $F\ (2, 46) = 13.85$, to be significant with a $p < .01$, with no Trials × Weeks interaction.

The significant decrement in mean GSR magnitude from Week 1 to Week 2, $p < .01$, demonstrates long-term retention of habituation despite a change in the stimulus modality. Figure 1b shows that the mean GSR magnitude on Trial 1 was invariant across weeks.

Habituation was considered to be specific to the modality of the stimulus to the extent that the Same group had a smaller GSR magnitude than the Changed group, particularly on Week 2. To insure that performance differences between these groups during original learning (Week 1) did not contaminate the comparison, the GSR magnitude/trial for each S on Weeks 2 and 3 was subtracted from the mean GSR magnitude during Week 1 for that same S.

Analysis of the variation between the deviation scores/S/trial/wk showed the effects of weeks, $F\ (1, 46) = 4.77$, $p < .05$, trials, $F\ (19, 874) = 28.56$, $p < .01$, and Trials × Groups, $F\ (19, 874) = 3.66$, $p < .01$, to be reliable.

A priori interest in the pattern of differences between groups over weeks and trials justified an examination of the Groups × Weeks × Trials interaction. Analysis of the early trials on Weeks 2 and 3 with the Tukey *a* procedure indicated that the Same Ss responded less than the Changed Ss on Trials 1, 2, and 3 on Week 2 and only on Trials 1 and 2 for Week 3, $p < .05$. Thus, the conclusion that habituation is a modality-specific phenomenon is warranted for the performance on only the first few trials of each week.

The SRL was defined as the skin resistance in ohms at the time of stimulus occurrence. Analysis of the SRL/trial/S on Week 1 for the Same and Changed groups indicated no difference between groups, $F\ (1, 46) = .03$.

Figure 2 shows the mean SRL/trial/wk for each group. Analysis of the SRL/trial/S/wk/group showed trials, $F\ (19, 874) = 13.53$, $p < .01$, Groups × Trials, $F\ (19, 874) = 2.84$, $p < .01$, and Groups × Weeks × Trials, $F\ (19, 874) = 1.86$, $p < .05$, to be reliable. Analysis of the third-order interaction with the Tukey *a* procedure indicated that while the SRLs of the Changed and Same groups were similar at the beginning of the Week 2 session, by the end of that session the SRL of the Changed group was greater than that of the Same group, $p < .01$. Figure 2 shows that this difference persisted through Week 3.

Thus, to the extent that SRL is a measure of arousal (Malmo, 1962), Ss with prior experience of novelty were less aroused. The inverse relationship between novelty and arousal suggests that the large GSR on Trial 1 of Week 2 and 3 for the Changed Ss was not an artifact of an increment in arousal produced by novel stimulation (Thompson & Spencer, 1966). In fact, the SRL for the Changed Ss was significantly higher than that of the Same Ss just *after* novel stimulation (Trial 2, Week 3), $p < .01$.

REFERENCES

Kimmel, H. D., & Goldstein, A. J. Retention of habituation of the GSR to visual and auditory stimulation. *Journal of Experimental Psychology*, 1967, **73**, 401–404.

Malmo, R. B. Activation. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology*. New York: Basic Books, 1962.

Thompson, R. F., & Spencer, W. A. Habituation: A model phenomenon for the study of neuronal substrates. *Psychological Review*, 1966, **73**, 16–43.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received January 31, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 393–395

INTERACTIVE EFFECTS ON REACTION TIME OF PREPARATORY INTERVAL LENGTH AND PREPARATORY INTERVAL FREQUENCY

ALFRED A. BAUMEISTER[1] AND CHARLES E. JOUBERT

University of Alabama

Simple reaction times (RTs) were found to be dependent on the interactive effects of length of the preparatory interval (PI) and the distribution of PIs within a series. Four PIs (2, 4, 8, and 16 sec.) were presented under four distributions: rectangular, symmetrical, skewed left, and skewed right. For the first three distributions, RTs following the shortest PI were particularly slow. However, in the case of the distribution containing a relatively high proportion of short intervals (skewed right), no differences in RTs were observed in relation to length of the interval. An additional result indicated that performance under the long PIs was detrimentally affected over trial blocks regardless of the distribution.

It is well known that within certain limits, the length of the foreperiod affects reaction time (RT). Moreover, when the foreperiods are variable from trial to trial, a contextual factor is introduced in which performance on a given trial is influenced by the foreperiods on preceding trials, particularly the one most immediately antecedent (Karlin, 1959; Klemmer, 1956). This phenomenon has been called the previous preparatory interval (PPI) effect. The PPI effect tends to be most marked in the situation where a trial defined by a short foreperiod follows a trial associated with a long foreperiod. The *S* behaves as if he has developed an expectancy for a long interval and is caught unprepared when the interval turns out to be short.

One way that has been proposed to manipulate this expectancy is to vary the relative frequency of the various preparatory intervals (PI). For example, Zahn and Rosenthal (1966), using only two PIs at a time, found that when the series included a relatively small number of brief PIs, RTs were particularly slow on trials having the short intervals. In other words, a PI Frequency × PI interaction was observed. Karlin (1966) used three distributions of PIs: leptokurtic, rectangular, and bimodal. The major finding was that readiness to respond was dependent on the probability of occurrence of an interval of a given length. Karlin's study is somewhat difficult to interpret in that different ranges of PIs and different numbers of trials were given under the various distributions. In addition, the ranges of PIs were short, the largest being 150–950 msec. It is not clear whether these findings can be generalized to an RT task involving a variable series of longer PIs.

The present study was undertaken to vary systematically the shape of the PI distribution (with respect to frequency) within a series. PI distributions were rectangular, skewed left, skewed right, and symmetrical. The range of PIs was 2–16 sec.

Method.—Sixty-three undergraduates enrolled in an introductory psychology class served as *S*s. Data from two *S*s were discarded due to equipment failure, and one *S* was excused for failure to follow instructions.

The equipment consisted of a telegraph key, a circular red light 1 in. in diameter mounted in a display unit, a Beltone audiometer with sound-attenuating earphones, an eight-bank timer, a Hunter Klockounter, and other programming equipment. The red light served as a warning signal, a 50-db., 1,000-cycle tone as the reaction signal. The experiment was conducted in a sound-dampened room under a low level of illumination. Controlling and recording equipment was located in an adjacent room. The warning signal appeared for 1 sec., followed by a variable blank interval of 1, 3, 7, or 15 sec. Preparatory intervals were thus 2, 4, 8, and 16 sec. Four numerically equal ($N = 15$) subgroups were constituted. Each subgroup was run under a different PI distribution over 100 trials: rectangular (R), symmetrical (S), skewed right (SR), and skewed left (SL). The total number of presentations of the various PIs for the four distributions is presented in Table 1. The 100 trials were presented in five blocks of 20, with each PI represented an appropriate number of times within a block depending on the distribution.

The *S*s were instructed to depress the telegraph key when the red light appeared and to hold the key down until the tone was presented. Emphasis was placed on making rapid responses. No reference was made to the length of the preparatory interval.

Results and discussion.—Obviously, when the PI distribution is anything but rectangular, the number of RTs vary from PI to PI. In analyzing the results, one may compute a mean based on all trials for a particular PI, in which case the number of observations contained in a measure will be different among the various PIs. Another procedure is to sample randomly, from the trials given at each PI, a number equal to the number given under the least frequent PI. Both procedures

---

[1] Requests for reprints should be sent to Alfred A. Baumeister, Department of Psychology, University of Alabama, University, Alabama 35486.

TABLE 1

Total Number of Presentations of the Various PIs for the Four PI Distributions

PI distribution	PI (in sec.)			
	2	4	8	16
R	25	25	25	25
S	10	40	40	10
SL	10	20	30	40
SR	40	30	20	10

were followed in the present study and the results obtained were nearly identical. Consequently, the results reported here are based on *all* 100 trials.

A 4 × 4 × 5 analysis of variance was conducted involving (*a*) type of PI distribution, (*b*) PI, and (*c*) trial blocks. A summary of the analysis is presented in Table 2. The main effects of PI and trial blocks were significant. In addition, trial blocks interacted with PI and type of distribution interacted with PI. This latter interaction is of major interest to the present investigation. These effects are graphically represented in Fig. 1 and 2.

An examination of Fig. 1 suggests that the inferiority sometimes observed under short PIs in an irregular series does not occur when the majority of the PIs in the distribution are relatively short (skewed right). One possibility is that *S*s "tuned in" to the statistical properties of the intervals and did not develop high expectancies for the long PIs. An examination was made of the contextual effects of the intervals for the rectangular distribution and the distributions of PIs skewed right. In the case of the former, the means for the longest PPI and the shortest PPI were .37 and .32 sec., respectively. For the skewed-to-the-right distribution, the corresponding means were .34 and .29 sec. An analysis of these data revealed that RTs to a short PI preceded by a long PI were slow ($p < .05$) relative to those observed when the PPI was also brief. Furthermore, this effect was approximately the same for the two distributions. Thus, the interaction graphed in Fig. 1 may have resulted simply from the relatively few occasions for a long PPI in the distribution skewed right.

TABLE 2

Summary of Analysis of Variance

Source of variance	*df*	*MS*	*F*
Between *S*s			
Treatments (A)	3	28227.78	<1
S × A	56	45757.81	
Within *S*s			
PI (B)	3	80988.81	39.04**
A × C	9	8857.54	4.27**
S × B × A	168	2074.72	
Trial blocks (C)	4	18956.19	8.82*
A × C	12	1276.37	<1
S × C × A	224	2148.29	
B × C	12	3843.25	3.68**
A × B × C	36	1105.94	1.06
S × B × C × A	672	1044.56	

* $p < .01$.
** $p < .001$.

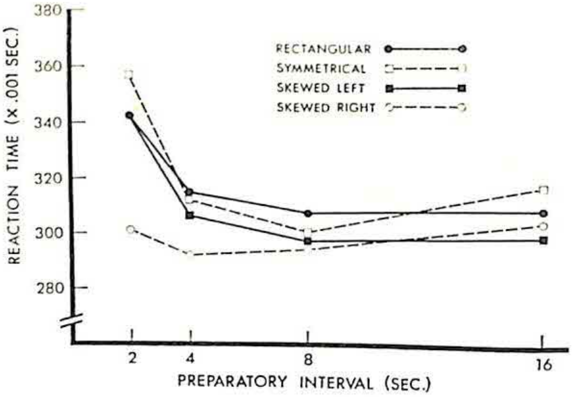


Fig. 1. Mean RTs at each PI for four types of PI distribution.

It should be noted also that the shape of the foreperiod distribution did not influence performance under the longer PIs. In other words, the ability to maintain a set for the long intervals was not dependent on their frequency. On any given trial, *S* is given more information about the occurrence of a long interval than a short one. That is, once *S* has assimilated the fact that the intervals will vary in length, he can "check off" the probability of a longer interval by eliminating the shorter PIs as they fail to occur on a given trial. If *S* is set for a short interval and it does not occur, then he "knows" that it must be a longer one and therefore sets himself again.

In view of these findings as well as the failure to find a significant Blocks × Treatment interaction, it is not entirely compelling to attribute the PPI effect to "expectancy." Possibly, a long PPI produces a temporary adaptation effect in which the duration of shorter PIs is overestimated by *S*. A similar explanation has been proposed by Zahn and Rosenthal (1966).

The Trial Blocks × PI interaction may have arisen from a deterioration in *S*'s ability to maintain a preparatory set for the long intervals. Figure 2 indicates that performance under the 2-sec. PI did not change markedly from Trial Blocks 1

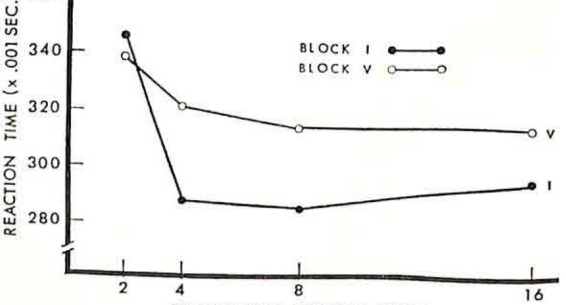


Fig. 2. Mean RTs at each PI for Trial Block 1 and Trial Block 5.

to 5. Another possibility is that the presence of short intervals in the series somehow interfered with the ability to maintain a set over longer periods. However, this alternative explanation suffers from the absence of a statistically significant three-way interaction. In other words, one might have expected the deterioration in performance under the long PIs to have been more evident when there was a relatively high proportion of short PIs.

REFERENCES

Karlin, L. Reaction time as a function of foreperiod duration and variability. *Journal of Experimental Psychology*, 1959, **58**, 185–191.

Karlin, L. Development of readiness to respond during short foreperiods. *Journal of Experimental Psychology*, 1966, **72**, 505–509.

Klemmer, E. T. Time uncertainty in simple reaction time. *Journal of Experimental Psychology*, 1956, **51**, 179–184.

Zahn, T. P., & Rosenthal, D. Simple reaction time as a function of the relative frequency of the preparatory interval. *Journal of Experimental Psychology*, 1966, **72**, 15–19.

(Received February 27, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 395–397

COMPARISON OF THE EFFECT OF AUDITORY VERSUS VISUAL STIMULATION ON INFORMATION CAPACITY OF DISCRETE MOTOR RESPONSES

W. W. BREEN, M. J. DE HAEMER, AND G. K. POOCK[1]

Naval Postgraduate School, Monterey, California

The effects of auditory and visual stimulation on reaction time (RT) and movement time (MT) for discrete motor responses were compared with the effects of visual stimulation studied by previous investigators. Under audio stimulation, the correlation between MT and information processed was .92. Under visual stimulation, the correlation between MT and information processed was .97. Although the amount of information processed affected MTs, RT was not affected by this variable. At the same time, however, type of stimuli did affect RT. RT to visual stimuli was faster than RT to audio stimuli. Relative independence of RT and MT under audio stimuli further supports the independent nature of perceptual and motor processes as asserted by previous investigators using visual stimuli in similar experiments.

Fitts and Peterson (1964) investigated the relationship between the information contained in a motor response, the movement time (MT) to accomplish the response, and the reaction time (RT) for the decision process associated with the response when a visual stimulus was used. They defined an index of difficulty (ID) as

$$\mathrm{ID} = \log_2 \left(\frac{2A}{W}\right),$$

where A is the movement amplitude between a starting button and a target and W is the target width. Using two targets of width W and a distance between them equal to $2A$, their *S*s held a stylus at a center starting button between the two targets and moved the stylus to the right or left target on presentation of the proper visual stimulus. RT and MT were measured, with subsequent analysis showing a linear relationship $\mathrm{MT} = a + b(\mathrm{ID})$ and a relative independence between RT and ID.

These results appeared applicable to the current authors' problem of trying to predict and improve overall RTs involving various basic movements in certain military systems. However, the current problem included prediction of RTs to auditory as well as visual stimulation such as found in sonar systems. Fitts and Peterson's (1964) experiment was therefore duplicated with the addition of auditory stimulation.

Method.—Six male military officers ranging from 28 to 31 yr. of age served as *S*s.

The apparatus was designed to be as nearly as possible like the apparatus detailed in Fitts and Peterson (1964). Two identical targets were located at equal distances from the starting plate. The *S*s held a lightweight metal stylus on the starting plate. In the audio stimulus trials, *S*s moved the stylus as rapidly as possible to the left

---

[1] Requests for reprints should be sent to Gary K. Poock, Department of Operations Analysis, Naval Postgraduate School, Monterey, California 93940.

TABLE 1

Amplitude and Target Characteristics Associated with Indexes of Difficulty

Amplitude (A)	Target width (W)	Index of difficulty (ID)
3	1	2.58
6	1	3.58
12	1	4.58
12	.5	5.58
12	.25	6.58
12	.125	7.58

Note.—Measures of amplitude and target width are given in inches.

target on presentation of a low-pitched tone (250 Hz.) or to the right target when presented with a high-pitched tone (2,000 Hz.). The tones were presented through earphones. In the visual stimulus trials, *S*s observed two lights and moved the stylus in the direction indicated by the light stimulus in any given trial. RT was measured from the onset of a stimulus to the time the stylus was lifted off the starting plate. MT was measured from the end point of the RT period to the time the stylus touched the target. Times were measured to the nearest millisecond.

For comparison with the results of Fitts and Peterson (1964), IDs identical to theirs were used. Each ID was used under each stimulus condition. The target characteristics are shown in Table 1.

Each *S* performed 10 trials under each ID and stimulus condition after performing several practice trials for familiarization. Times for practice trials were not recorded. Subsequent analysis verified previous experience that no learning occurred during the experiment, and each measure

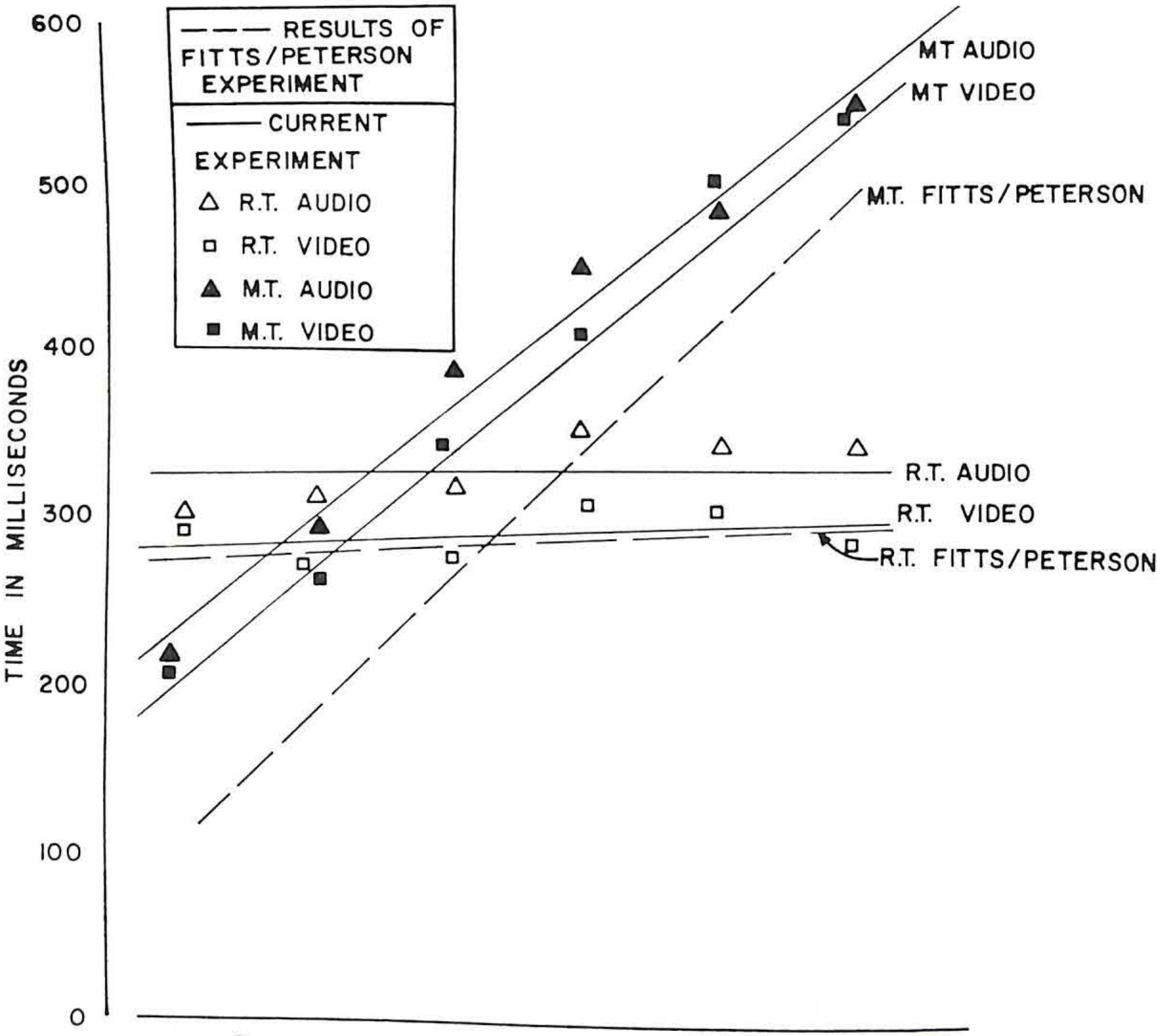


FIG. 1. Index of difficulty vs. time.

was therefore considered independent. For any given 10 trials, *S*s always performed under one ID condition and one stimulus condition. The *S*s were instructed to move the stylus to the appropriate target when the stimulus occurred. ID conditions, stimulus type, and direction of movement were randomized across all *S*s. Signaled direction was made to occur randomly but with equal number in each direction for each ID condition. Trials in which errors (misses of the target) occurred were infrequent and disregarded. Stimuli for given trials were presented at the rate of approximately 5/min.

Results and discussion.—A two-way analysis of variance was performed on RT and MT for effects of ID and stimulus. Stimulus had a significant effect on RT, $F\ (1, 708) = 82.64$, $p < .001$. Mean RT for visual stimulus was 33 msec. shorter than for auditory stimulus, as shown in Fig. 1. ID did not have a significant effect on RT, $F\ (5, 708) = 1.90$, $p > .05$. The correlations between RT and ID values were .202 (auditory) and .413 (visual).

Changes in ID had a significant effect on MT, $F\ (5, 708) = 35.30$, $p < .001$, but stimulus was not significant in determining MT, $F\ (1, 708) = 1.95$, $p > .10$. The correlations between ID and MT were .918 (auditory) and .973 (visual). Overall, when considering the amount of variance accounted for, ID was the most influential variable in the present experiment.

The regression equations for predicting MT are given as follows: (*a*) $\text{MT}_{\text{visual}} = 71\ \text{ID} + 16$ (msec.), (*b*) $\text{MT}_{\text{audio}} = 67\ \text{ID} + 59$ (msec.), (*c*) $\text{MT}_{(\text{F and P})} = 74\ \text{ID} - 70$ (msec.). Equation *c* is the one from the earlier Fitts and Peterson (F and P; 1964) experiment relating MT to ID under visual stimulus conditions.

A major goal of the current experiment was to validate, for an audio stimulus, the relationship between MT and ID which was developed by Fitts and Peterson (1964) for visual stimulus. All three of the regression equations relating MT to ID have essentially the same slope. The difference in intercept constants could be the result of slight differences in experimental conditions and may also be due to the age of *S*s used. The *S*s in the current experiment showed no learning effects and were from 5 to 10 yr. older than those used by Fitts and Peterson. The difference in the intercept constants of the MT regression equations between Fitts and Peterson's Exp. II and the current experiment lends support to the empirical evidence that age may affect psychomotor processes of this type, as suggested by Birren, Butler, Greenhouse, Sokoloff, and Yarrow (no date) and Welford (1958).

In conclusion, the equation $\text{MT} = a + b(\text{ID})$ first demonstrated by Fitts and Peterson has been reconfirmed for *both* audio and visual stimulus. The current experiment has revealed that only the type of stimulus significantly affected RT; and ID alone significantly affected MT. Relatively, ID was a much more influential variable.

REFERENCES

BIRREN, J. E., BUTLER, R. N., GREENHOUSE, S. W., SOKOLOFF, L., & YARROW, M. R. (Eds.) Human aging. United States Public Health Service Publication No. 986, United States Department of Health, Education, and Welfare.

FITTS, P. M., & PETERSON, J. R. Information capacity of discrete motor responses. *Journal of Experimental Psychology*, 1964, **67**, 103–112.

WELFORD, A. T. Ageing and human skill. Oxford: Oxford University Press, 1958.

(Received March 6, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 398–400

SERIAL ACQUISITION AS A FUNCTION OF STAGE OF LEARNING[1]

M. AHMED AND JAMES F. VOSS[2]

University of Pittsburgh

Following a previous study, *S*s were interrupted at one of five stages of serial learning ranging from a 0/12 criterion (one trial) to a criterion of 12/12 + 10 extra trials. Upon interruption, *S*s first tried to recall the list items and then were presented each item and asked to place the item along a spatial dimension representing the serial list. In addition, two control conditions were employed in which *S*s either placed the items along the dimension with no information or with complete information regarding the list locus of the items. The results indicated that one trial reduces the average deviation of placement by approximately one-half and that placement performance tends to improve during acquisition. The results were interpreted as supporting the view that the serial position curve is attributable to *S*'s knowledge of the dimensional characteristics of the serial task.

In a previously reported experiment (Voss, 1969), acquisition of a 12-unit serial list was studied as a function of stage of learning. The *S*s were interrupted at one of five criteria: 0/12 (one presentation), 4/12, 8/12, 12/12, or 12/12 + 10 extra trials. One of three interruption tasks was then employed: recalling as many items as possible without reference to serial position, stating which item of the list followed each item when the items were presented in counterbalanced orders, and stating which item occurred at a particular serial position when the digits 1–12 were presented. The results provided little evidence supporting the serial position and chaining hypotheses, but were interpreted in terms of a view stating that *S*s acquire a serial list by learning to place items along a dimension (Slamecka, 1964; Voss, 1968).

The present paper reports the results of an experiment designed to supplement the previous study by use of a different interruption task. Specifically, after reaching the respective criterion, *S*s were asked to try to recall the items and then were presented with a horizontal line representing the serial list. The *S*s subsequently were presented each item and asked to draw an arrow at the point where the item occurred, with performance measured in terms of deviation from correct placement. It was hypothesized that deviation scores decrease as a function of trials, that deviations of recalled items are less than deviations of nonrecalled items, and that the deviation of items correct on the trial preceding the interruption is less than the deviation of items not correct.

Method.—The same 12-unit serial list was employed as that of the previous experiment (Voss, 1969). Five stages of learning were used: 0/12 (1 trial), 4/12, 8/12, 12/12, and 12/12 + 10 trials. Upon interruption, *S*s were given two tasks. The *S*s were provided a sheet of paper and asked to write all the items they were able to without regard to serial position. Following recall, *S*s were given a sheet of paper with a horizontal line and were instructed to draw an arrow on the line at the point that each item occurred. Each end of the line was marked with a short vertical line that represented the first and last items. The line was 8¼ in., with each ¾ in. representing the correct placement of the item. The items were presented to each of 12 *S*s in an order determined by a 12 × 12 Latin square. A new sheet of paper was provided for the placement of each item. Twelve *S*s were run in each of the five degree-of-learning conditions; one-half of the *S*s were presented the recalled items for placement prior to presentation of the nonrecalled items, whereas the remaining one-half of the *S*s were presented the nonrecalled items first.

In addition to the five experimental conditions, two control conditions were employed. In Cond. C_1, *S*s were presented the items and asked to place them along the line without having any exposure to the list; in Cond. C_2, *S*s placed the items with the specific serial position of each item given. The former condition was employed as a control for item placement with no serial list information; the latter condition was employed as a control for assessing serial placement performance with complete information given.

The *N* per condition was 12, or a total *N* of 84. Other conditions were comparable to those of the previously reported experiment (Voss, 1969), with the exception that *S*s of Cond. C_2 were run after the other six conditions.

Results and discussion.—Deviations from correct placement were measured in terms of ⅛ in. Table 1 presents the average deviation scores for the experimental conditions. Three of the comparisons are especially noteworthy: (*a*) The difference between the C_1 and 0/12 conditions shows that in one trial the deviations were reduced by approximately 50%, thus indicating that one presentation of the list provides substantial information about the loci of list items; (*b*) as degree of learning increased, the average deviation decreased, with an

---

[1] This research was supported by National Institute of Child Growth and Human Development Grant HD00957-05.

[2] Requests for reprints should be sent to James F. Voss, Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

TABLE 1

AVERAGE DEVIATION ($\frac{1}{8}$ IN.) PER ITEM FOR EXPERIMENTAL AND CONTROL CONDITIONS

Cond.	AD		
	Placement	Recalled items	Nonrecalled items
C_1	51.1	—	—
0/12	25.2	21.4	30.6
4/12	11.6	11.4	13.2
8/12	15.7	15.3	18.2
12/12	10.8	—[a]	—[a]
12/12 + 10	6.7	—[a]	—[a]
C_2	5.5	—	—

[a] Statistic was omitted because of insufficient data points.

TABLE 2

AVERAGE DEVIATION OF COND. 4/12 AND 8/12 ITEMS FOR CORRECT AND INCORRECT CRITERION TRIAL AND RECALL PERFORMANCE

Criterion trial	Cond. 4/12		Cond. 8/12	
	Recalled	Not recalled	Recalled	Not recalled
Correct	9.4	13.6	13.5	21.3
Incorrect	13.5	12.1	22.6	16.8

inversion at the 4/12 and 8/12 conditions; (*c*) the decrease in average deviation from the 12/12 to the 12/12+ criteria indicates that the deviation measure is sensitive to overlearning effects. In addition, it should be noted that the average deviation of the 12/12 + 10 condition is only slightly greater than that of the C_2 control condition, thus suggesting that greater overlearning of the list would produce only a small gain in item placement performance.

An analysis of variance indicated a significant effect of condition, $F\ (6, 77) = 68.07$, $p < .01$, and a subsequent Duncan range test revealed that Cond. C_1 differed significantly from all other conditions, as did Cond. 0/12. The remaining five conditions had three overlapping ranges of significance. An analysis of variance conducted on the five experimental conditions revealed that condition is significant, $F\ (4, 50) = 10.86$, $p < .01$, and that the recall–nonrecall variation in item placement and the Condition × Recall–Nonrecall Order sources of variation are not significant, $F < 1.00$ in both cases. These data thus indicate that average placement deviation does not vary as a function of whether the recalled items or the nonrecalled items are placed first.

The last two columns of Table 1 present the average deviation scores for the items recalled and not recalled. An analysis of variance revealed that the average deviation of recalled items was smaller than that of nonrecalled items at the 0/12 condition, $F\ (1, 144) = 4.01$, $p < .05$, but at the 4/12 and 8/12 conditions, the average deviation of recalled and nonrecalled items is not significant, $F < 1.00$ in both cases. Thus, contrary to hypothesis, recalled items did not yield significantly smaller deviations than items not recalled, except in the 0/12 condition.

Table 2 presents the average deviation for the 4/12 and 8/12 conditions, tabulated on the basis of whether an item was correctly anticipated on the respective 4/12 or 8/12 criterion trial and whether or not it was recalled correctly. The data indicate that the best performance in both conditions occurred when an item was correct on the criterion trial and correctly recalled. One reason for the surprisingly small deviation of items neither correctly anticipated nor recalled may be that *S*s, when presented these items, realized they did not know their correct locus and tended to place them in a part of the list they knew they did not know, i.e., the middle.

Figure 1 presents the average deviation for each condition as a function of serial position. Three findings are of note: (*a*) Average deviation tends to increase over serial position for the 0/12 condition, thus suggesting that in one trial, *S*s show a greater gain in placement for items in the first-half positions of the serial list; (*b*) the remaining curves tend to approach the shape of the serial position curve as acquisition continues; (*c*) the discrepancy between Cond. 12/12 + 10 and C_2 tends to take place at the serial positions just past

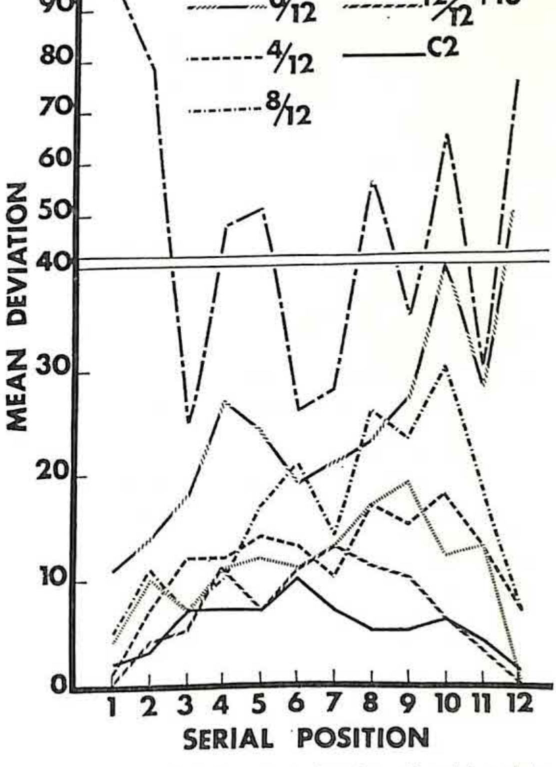


FIG. 1. Average deviation as a function of serial position for the item placement of the seven conditions.

the middle of the list. It also should be mentioned that serial position placement of items recalled tended to follow the serial position curve to a greater degree than items not recalled (data not presented).

The results of the present experiment supplement the previous findings (Voss, 1969) by showing that as *S*s learn a serial list, accuracy of placement of items increases along a spatial dimension representing the list. Furthermore, it is of interest to note that Cond. C_1 and C_2 yielded an irregular curve and a bow-shaped curve, respectively, without the list per se ever occurring in either position. In other words, the serial position curve was generated without the typical repeated presentation of the serial list. Finally, it may be mentioned that the present data support a positional hypothesis, which affirms that *S*s tend to acquire a list by a process of placing the items along a dimension; the test dimension in the present study, however, was spatial, whereas the typical serial tasks intrinsically involve a temporal dimension or, possibly, a temporal dimension that *S* may transform and encode into a spatial dimension.

REFERENCES

SLAMECKA, N. J. An inquiry into the doctrine of remote associations. *Psychological Review*, 1964, **71**, 61–76.

VOSS, J. F. Serial acquisition as a function of number of successively occurring list items. *Journal of Experimental Psychology*, 1968, **78**, 456–562.

VOSS, J. F. Serial acquisition as a function of stage of learning. *Journal of Experimental Psychology*, 1969, **79**, 220–225.

(Received March 7, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 400–402

WORDS AS FEATURE COMPLEXES: FALSE RECOGNITION OF ANTONYMS AND SYNONYMS[1]

SAMUEL FILLENBAUM[2]

University of North Carolina at Chapel Hill

For each item in a 240-word list that was orally presented, *S*s had to indicate whether or not it had occurred previously. There were significantly more false recognition errors for (*a*) antonyms and (*b*) synonyms of preceding words than for control items matched with *a* and *b* for strength of associative relation to preceding words. There was no difference between the antonym and synonym conditions. These findings, which extend some earlier work of Anisfeld and Knapp on false recognition for synonyms, were taken as evidence for the view that a word is understood and stored in memory as a complex of features or attributes, systematic errors occurring when information on some feature(s) is lost.

It has been suggested that a word is understood and may be represented in memory as an ordered list of features or attributes (see, e.g., Bower, 1967). Such an approach not only has the virtue of economy, since inevitably the number of elementary meaning processes (markers) out of which the meaning of words is constructed will be vastly less than the number of different words, but also allows us to conceive of the meaning relations among words in terms of overlap in their component markers.

The study of systematic properties of errors in recall may tell us something about the way in which items are understood and stored. In particular, given an interest in problems of meaning, one may examine recognition errors in a situation where words stand in some specifiable semantic relation to each other. This was done by Anisfeld and Knapp (1968), using a continuous recognition technique. They found that synonyms of preceding words were significantly more often falsely recognized than were control words and also demonstrated that such findings could not be regarded simply as artifacts of testing, but stemmed from the initial coding of the words. These data are clearly consistent with the aforementioned view, viz., that a word is stored as a feature complex. If some of these features are disregarded or lost, then the synonym of a word, which has most of the same attribute values as the word itself, may well be falsely recognized as having occurred previously. The authors interpret their data in precisely these terms. However, while consistent

---

[1] This study was carried out with partial support from United States Public Health Service Grant MH10006 from the National Institute of Mental Health. The author is greatly indebted to Sandra Funk for her aid in all phases of the work.

[2] Requests for reprints should be sent to Samuel Fillenbaum, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina 27514.

with such a view, these data on synonym confusion do not require it. Consider that just as one may store the gist of a sentence, so one may store the sense of a word, i.e., note the concept embodied or realized in that word, and that concept need not be *analyzed* as a feature complex.[3] Now, when a new word is encountered, one may check to determine whether or not the concept that it represents has occurred previously, and if the answer is "yes," one may say that the word has occurred previously. Given that synonyms embody roughly the same concept, one could thus readily account for the findings of Anisfeld and Knapp. Obviously, what is needed is a task that may permit us to discriminate between these two interpretations. A situation where one tests for false recognition of antonyms may do just this. On the feature complex view, one can account for the closeness of opposites by noting that for any antonymous pair, all semantic markers save one will be identical, the two words contrasting in their value on that feature or attribute. If information on the critical contrasting feature is lost, then an antonym may well be falsely recognized as having occurred previously. On the more global interpretation previously noted, the two members of an antonym pair in fact represent opposite concepts, and one should therefore expect little confusion between them. These considerations governed the choice of the antonymy relation in the present study, which may be seen essentially as a replication and extension of the work of Anisfeld and Knapp by means of an assessment of false recognition errors both for antonyms and synonyms.

Method.—Since it has been shown previously that associatively related items may be falsely recognized (Anisfeld & Knapp, 1968), and since it is also well known that antonymous items are often highly related in associative terms (see, e.g., Deese, 1965), it will not do simply to take antonymous terms (or synonymous ones) and test for false recognition. Rather, what is required is that the degree of associative relation be controlled while the conceptual relation among item pairs is systematically varied.

A list was constructed so as to contain preceding (P) words and following (F) words (plus some filler items). The critical relation was that between P and F words. There were three possible relations: (*a*) An F word might be a synonym (syn) of a P word; (*b*) an F word might be an antonym (ant) of a P word; or, (*c*) an F word might be neither synonym nor antonym (control). The "Atlas of Normative Free Association Data" compiled by Shapiro and Palermo (1968), which lists the associative probabilities of primary responses obtained in a variety of studies, was searched for words with primaries in the ranges .30–.39, .40–.49, and .50–.59. Control item pairs were selected from these ranges, as well as pairs satisfying the synonymy condition and the antonymy condition. In each case there were 7 pairs from the .30–.39 range, 6 from the .40–.49 range, and 7 from the .50–.59 range, for a total of 20 item pairs in each condition. Items for the three conditions were matched for strength of associative relation. The mean values for the synonym, antonym, and control conditions were .438, .437, and .440, respectively. Since we also wished to examine possible misrecognition effects for antonyms only weakly related in associative terms, some of the association lists cited by Shapiro and Palermo were searched for word pairs that were antonymous but where the associative probability was less than .10 (ant–low-ass); 10 such pairs were selected.

In the list there were 70 different P words and 70 different F words. Each P word appeared twice in different positions before its F word appeared, the two tokens of each P word being within 20–50 positions of each other, and each F word appeared within 20–50 positions after the second token of its P word. This accounts for 210 words. In addition, there were 30 filler words. These occurred in the first 10 positions on the list and near and at the end of the list.

The list of 240 words was tape-recorded, the words being presented at 5-sec. intervals. Each word was recorded twice in immediate succession so as to make it more likely that it would be heard correctly. The tape was played to *S*s in groups of about 10. For each word, *S* had to indicate on his response sheet by marking + or − whether or not it had occurred previously on the list. The *S*s were instructed to guess when in doubt and asked not to omit any items.

There were 79 *S*s, students meeting a course requirement for introductory psychology.

Results and discussion.—For each *S*, the number of false recognitions was noted for each of the four classes of items, and analyses were performed on the error proportions, with each *S* as his own control by serving in all conditions of a repeated-measures design.

While the error rate in the Anisfeld and Knapp (1968) study was 5–6%, it was 15–16% in the present study. This difference is reasonable enough if it is recalled that in the earlier study each P word was presented three times for 10 sec. each, while in the present study each P word was presented only twice for 5 sec. each.[4]

The mean proportions of false recognition errors for the syn, ant, control, and ant–low-ass conditions were .170, .193, .117, and .136. An analysis of variance on the error proportions yielded a highly significant *S* effect, F (78, 234) = 7.54, $p < .001$, and, more to the point, a highly significant conditions effect, $F\ (3, 234) = 14.05$, $p < .001$. Contrasting the various means by use of t tests, it was found that both the syn and ant conditions yielded significantly more false recognitions than the control condition, t's $(78) = 5.12$ and 6.63, respectively, $p < .001$ in each case, and that the difference between the former conditions was not significant. Nor was the difference between the ant–low-ass and the control condition significant. There was a difference between syn and ant–low-ass, $t\ (78) = 2.48$, $p < .05$, which is uninterpretable since both the nature of the conceptual relation and the strength of the associative relation differ between these conditions, and finally there was a difference between ant and ant–low-ass, $t\ (78) = 3.80$, $p < .01$, suggesting that the strength of the associative relation between items may be a determinant of false recognitions, a result consistent with earlier findings.

---

[3] Data compatible with this sort of characterization may be found in some studies of bilingualism where, with regard to later recall, it does not seem to matter very much whether a word has occurred n times in one language or whether two translation equivalents of it have each occurred with a frequency of $n/2$ (Kolers, 1966).

[4] The relatively low rate of false alarms indicates that verbatim memory is good in the present situation and should serve to remind us that there may be other components of word coding and storage than those discussed here, e.g., storage may also take place in terms of the phonological properties of words.

The fact that there is no difference in proportion of false recognition errors between the syn condition ($p = .170$) and the ant condition ($p = .193$) argues for the view that storage is in terms of some feature or attribute complex; otherwise, e.g., on a more global "concept storage" view, one would surely expect the former condition to yield more errors than the latter. The fact that there is no difference between the ant–low-ass condition ($p = .136$) and the control condition ($p = .117$) indicates again the importance of the semantic or conceptual relation between items, for, as noted earlier, the control condition is one with substantial associative relation between items (mean associative probability $= .440$). Thus the fact that items are opposites more than balances the much lower associative relation that holds in the ant–low-ass condition. Finally, and perhaps most important, both the syn and the ant conditions yielded significantly more false recognition errors than the control condition with which these conditions were matched in terms of strength of associative relation. All these findings are mutually compatible and consistent with the hypothesis that a word is coded and stored in memory as a feature or attribute complex.

REFERENCES

Anisfeld, M., & Knapp, M. Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology*, 1968, **77**, 171–179.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

Deese, J. *The structure of associations in language and thought*. Baltimore: Johns Hopkins University Press, 1965.

Kolers, P. A. Interlingual facilitation in short term memory. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 314–319.

Shapiro, S. I., & Palermo, D. S. An atlas of normative free association data. *Psychonomic Monograph Supplements*, 1968, **2**, 219–250.

(Received March 28, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 2, 403–404

NOISE, TASK DIFFICULTY, AND STROOP COLOR-WORD PERFORMANCE[1]

B. KENT HOUSTON[2]

University of Texas at Austin

Sixty-four college students performed either a high- or low-difficulty Stroop color-word (CW) task or high- or low-difficulty color-name (CN) task in a condition of either quiet (Q) or continuous noise (N). It was predicted that N would facilitate performance on the CW tasks, but not performance on high-difficulty tasks generally. Both predictions were supported. Pulse rate data taken immediately after Q or N did not indicate that N had produced physiological arousal. The results are discussed in terms of an interaction between the inhibitory process involved in ignoring N and the inhibitory process of ignoring the conflicting cues in CW.

Houston and Jones (1967) found that the presence of continuous, variable noise which *S* was instructed to ignore facilitated performance on a task which required him to inhibit responses to interfering cues (viz., the Stroop Color-Word Interference Test), but did not do so on a task which did not involve inhibition (viz., a color-naming task). In the Stroop color-word test (CW) the test items are the names of colors, each of which is printed in another color of ink; e.g., the word "red" might be printed in blue ink. The *S* is asked to name the color in which each word is printed, and in order to be successful at this task, *S* must inhibit responding to the word itself. A task similar to CW but which does not require inhibition of a prominent response is the color-naming task (CN). In CN, *S*s are asked to name the color of ink in which nonword stimuli are printed (e.g., asterisks, letters, color bars, etc.).

The results of the Houston and Jones (1967) study were interpreted as being due to an interaction between inhibitory processes. It was hypothesized that inhibiting a response to one source of extraneous stimuli (e.g., the noises) facilitates inhibiting response to another source of extraneous stimuli (e.g., ignoring the conflicting color names in CW).

There were two purposes of the present study. First, since in the Houston and Jones (1967) study significantly more time was required to complete CW than CN, $F\ (1,\ 41) = 421.07\ p < .001$, it is possible that noise facilitated performance on CW because it was more difficult than CN rather than because it required more inhibition. Therefore, in the present experiment, high- and low-difficulty versions of CW and CN were studied in order to assess whether task difficulty interacts with noise.

The second purpose was to obtain some measure of the extent to which the noise condition employed in these studies generated physiological arousal. This was done because noise has been reported to generate physiological arousal (cf Freeman, 1939), and it has been hypothesized tha a moderate increase in arousal decreases a person' attention to cues which are peripheral to task performance (Easterbrook, 1959). Therefore, if the presence of noise in the Houston and Jones (1967) study produced arousal, the arousal, independent of *S*'s trying to ignore the noise, may have decreased his attention to the interfering cues in CW, thereby facilitating performance on it. To provide information on this question, *S*s' pulse rates were taken immediately after performing in quiet or noise, and differences in pulse rates between the two conditions were compared.

Method.—The general procedure was to study performance on high- and low-difficulty versions of CW and CN in a continuous noise (N) condition in which *S*s had been instructed to ignore the sounds and in a quiet (Q) condition in which no noise was introduced.

In N, a variety of tape-recorded noises (e.g., trains, gibberish, electronic music, etc.) were played to *S*s through earphones at an average sound intensity of 78 db. The difficulty level of CW and CN were manipulated as follows. Three easily discriminated colors (red, yellow, and blue) were used for low-difficulty versions of CW and CN. Six less distinctly different colors (yellow, pink, orange, red, green, and blue) were employed in high-difficulty versions. In addition, the discriminability of these colors was decreased by decreasing the saturation and/or brightness of the six colors used. While in the first study, the items in CN consisted of five asterisks printed in the same color, a series of four H's were used in this study.

Sixty-four *S*s participated in the experiment and were randomly assigned to perform one of the four tasks: high- or low-difficulty CW or high- or low-difficulty CN, in one of the two treatment conditions, Q or N. Time to completion of each task was recorded and constituted the dependent

---

[1] This report is based on a doctoral dissertation submitted to Stanford University. The author is indebted to Eleanor Maccoby, Thomas K. Landauer, and P. James Geiwitz for helpful guidance and suggestions during the course of this study and to David Holmes and Ray Mulry for their constructive comments on the manuscript. This investigation was supported by United States Public Health Service Predoctoral Fellowship MH-15,757-04.

[2] Requests for reprints should be sent to B. Kent Houston, Department of Psychology, Mezes Hall, University of Texas, Austin, Texas 78712.

TABLE 1

MEAN TIME (IN SEC.) TO COMPLETE COLOR-WORD (CW) AND COLOR-NAME (CN) TASKS FOR TWO LEVELS OF DIFFICULTY UNDER THE EXPERIMENTAL CONDITIONS

Task	Low difficulty		High difficulty	
	Quiet	Noise	Quiet	Noise
CW	63.2	62.4	70.6	64.6
CN	38.0	44.5	49.1	56.8

measure. Immediately following task completion, *E* took *S*'s radial pulse manually, and the number of beats occuring during 15 sec. was recorded.

Results.—The means for the time taken to complete the experimental tasks, CW and CN, for the two levels of difficulty in the two treatment conditions, Q and N, are presented in Table 1. Upon inspection of the means, it may be seen that for both levels of difficulty, the time required to complete CW was less in N than Q and time to complete CN was greater in N than in Q. An analysis of variance indicated that this Treatments × Tasks interaction was significant, $F\ (1,\ 32) = 4.54$, $p < .05$, while the Treatments × Difficulty Level interaction was not significant, $F\ (1,\ 32) = .16$. The same analysis of variance indicated that the difficulty manipulation was successful in creating versions of the tasks that differed significantly in times to completion (high difficulty > low difficulty), $F\ (1,\ 32) = 11.21$, $p < .01$. A significant difference in times required to complete the tasks (CW > CN) was also indicated, $F\ (1,\ 32) = 54.12$, $p < .001$.

The mean numbers of pulse beats in 15 sec. after the two treatment conditions were as follows: Q = 19.31, N = 19.16. An analysis of variance indicated this difference in pulse rate between the two conditions was not significant.

Discussion.—The absence of a significant Treatment × Difficulty Level interaction indicates that performance on CW in Houston and Jones' (1967) study was not due to the fact that it was "just" more difficult than CN. The significant Treatments × Tasks interaction corroborates the findings of Houston and Jones' study that inhibition of attending to noise facilitates performance on a task requiring inhibition (CW), but does not do so for a similar task (CN) which does not require inhibition of the same magnitude.

The data on pulse rate failed to reveal an increase in physiological arousal following exposure to the noise condition. While the measure of arousal employed here is sufficient to detect large changes in arousal, it may not be sensitive to small changes which could have therefore gone undetected. Significant Conditions × Tasks interactions were obtained in this and Houston and Jones' study, yet no significant Tasks × Treatments interactions have been obtained in studies of arousal and CW performance in which arousal was generated by energizing drugs (Quarton & Talland, 1962). Therefore, it seems unlikely that a small increase in arousal if produced by N could account for the results of the present study and Houston and Jones' study. This consideration increases the credibility of the interpretation of the two studies in terms of an interaction between the inhibitory processes involved in not responding to the two sources of irrelevant stimuli.

REFERENCES

EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 1959, **66**, 183–201.

FREEMAN, G. L. Changes in tension-pattern and total energy expenditure during adaptation to "distracting" stimuli. *American Journal of Psychology*, 1939, **52**, 354–360.

HOUSTON, B. K., & JONES, T. M. Distraction and Stroop color-word performance. *Journal of Experimental Psychology*, 1967, **74**, 54–56.

QUARTON, G. C., & TALLAND, G. A. The effects of methamphetamine and pentobarbital on two measures of attention. *Psychopharmacologia*, 1962, **3**, 66–71.

(Received April 1, 1969)

Experimental and behavioral psychology require dependable classroom and laboratory instruments that will give years of reliable service—and that's the only kind of instrument we manufacture.

G Ralph Gerbrands Company
8 BECK ROAD, ARLINGTON, MASSACHUSETTS 02174 (617) 648-6415

Titles in Psychology from Macmillan

PSYCHOLOGICAL TESTING, Third Edition

By Anne Anastasi, Fordham University

Long praised as a book which is well-written, logically constructed, and extremely thorough in its coverage of both basic principles and the vast array of psychological tests available, *Psychological Testing* is now in a revised third edition. The text helps the student evaluate psychological tests of any type and interpret test results correctly by concentrating in this edition on the principles of test construction, the behavior the tests are designed to measure, and the field of available instruments. Multiple Choice Test Items are available gratis.

1968, 672 pages, $9.25

PRINCIPLES OF BEHAVIORAL ANALYSIS

By J. R. Millenson, Oxford University

This rigorous and systematic text is designed for use in first courses in psychology in which major concepts of functional behavioral analysis are built from the ground up. A unique feature of the book is its use of the *R* and *S* notation system as a kind of conceptualization and recording method. Elementary statistical concepts are developed within the framework of the data presented. Over 400 tables, line drawings, and diagrams are included. An Instructor's Manual is available gratis.

1967, 488 pages, $7.95

PERSONALITY: A Behavioral Analysis

By Robert W. Lundin, The University of the South

This book is a revision of the author's *Personality: An Experimental Approach.* It contains a significant amount of new material not available when the original volume was published in 1961. The title change reflects the addition of much contemporary research in the application of operant conditioning techniques to human behavior. Such topics as stuttering, human concept formation, autistic children, implicit response, and punishment are discussed. In addition, an entirely new chapter on behavior therapy has been included. An Instructor's Manual is available gratis.

1969, 464 pages, $7.95

EXPERIMENTAL SOCIAL PSYCHOLOGY

By Judson Mills, University of Missouri, Columbia

This book demonstrates the growing influence of the experimental method for the testing and development of ideas in social psychology. Each of the five sections, written by a contributor currently engaged in research in his designated area, represents a topic of major importance in the field. Significant experiments, most of them carried out within the last ten years, are discussed by: Dana Bramel, the State University of New York, Stony Brook ("Interpersonal Attraction, Hostility and Perception"); Ellen Berscheid, University of Minnesota, and Elaine Walster, University of Wisconsin ("Attitude Change"); Charles Kiesler, Yale University ("Group Pressure and Conformity"); Eugene Burnstein, University of Michigan ("Interdependence in Groups"); and Judson Mills, University of Missouri ("The Experimental Method").

1969, 466 pages, $9.95

Write to the Faculty Service Desk for examination copies.

THE MACMILLAN COMPANY
866 Third Avenue, New York, New York 10022
In Canada, write to Collier-Macmillan Canada, Ltd.,
1125B Leslie Street, Don Mills, Ontario

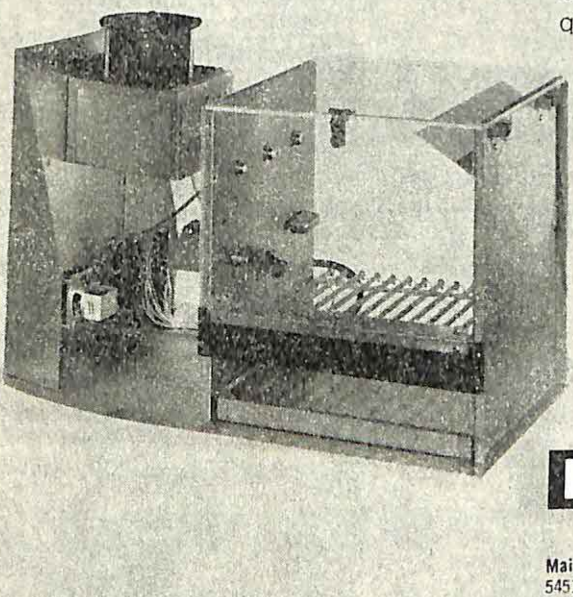
"Build a better rat test chamber and the world will beat a path to your door."

That's what we had in mind at BRS-Foringer when we talked with experimental psychologists and got their advice on re-designing and improving our complete line of environmental test chambers and support equipment.

Our new Environmental Equipment catalog shows the results—new rat and pigeon test chambers; our widely used primate chairs for Squirrel, Pigtail, Cebus and Rhesus monkeys. And BRS-Foringer backs them all with an array of environmental support equipment, including housings for all types of test chambers.

You will see a few unexpected innovations . . . a great many well thought out improvements designed to give you more flexibility and control in your laboratory.

Send for your copy of Environmental Equipment Catalog #469 by writing your nearest BRS-Foringer office. You'll discover that now you can get practically-priced environmental equipment to meet all your experiment requirements from one source.



BRS FORINGER®

A DIVISION OF TECH SERV, INC.

Main Office
5451 Holland Drive
Beltsville, Md. 20705
(301) 474-2900

Midwestern Office
1718 Sherman Avenue
Evanston, Ill. 60201
(312) 328-8883

Western Office
8949 Resada Blvd.,
Suite 203
Northridge, Calif. 91324
(213) 886-8010

The second family of curves (divider, DBG) indicates that the introduction of a divider produces an obvious reduction of slope for only *S*2, whereas there is an obvious reduction of slope for all *S*s viewing LBG sectors when the divider is introduced. Spatial changes in matched luminance values are nearly eliminated for *S*s 1 and 3 and are sharply reduced for *S*s 2 and 4. Under this condition, the mean matching luminance to Sector 6 is only 1.5 db greater than Sector 10.

Discussion

The important feature of the data presented in Fig. 1 is the regular spatial change in matched luminance value that occurs for Koffka's ring in the absence of any divider. The more tentative view based on the pilot investigation reported in the introduction is, therefore, confirmed.

It is instructive to compare the outcome of this experiment with that of Berman and Leibowitz (1965), who had their *S*s match one half of a transilluminated "figure 8" (DBG) to the other half (LGB). Their *S*s were instructed "not to try to match any one section of the figure with any other section, but to make the best overall match possible [p. 253]." Under these conditions, they report finding "considerable subjective contrast . . . in the absence of any dividing line between the two halves of the figure . . . [p. 256]." They suggest that the reason contrast effects are "probably not noticeable in the more familiar demonstrations of this phenomenon using paper stimuli," might be related to the presence of a "visible microstructure of the paper surface . . . [p. 256]." Of course, the pilot investigation reported earlier makes it clear that contrast is noticeable under analytic instructions when using paper stimuli.

One possible explanation for the usual failure to obtain reports of brightness differences when using paper examples of Koffka's ring is appealing in its simplicity. In the typical paper illustration, the *O* becomes "aware" that the gray ring is constructed of the same material throughout. When instructed to regard the ring as a whole, this "awareness" becomes the principal determinant of the verbal report. Stated otherwise, a cognitive report is substituted for a perceptual report. If this awareness is mediated by observing continuity in the visible microstructure of the paper surface, then the present hypothesis and that of Berman and Leibowitz (1965) are quite similar.

Certain features of the present experiment provide an interesting comparison to other data in the contrast literature. If, rather than making use of *S*s' attentional focus, the apparatus had been designed to permit the isolation and presentation of individual sectors, then observations made under the DBG condition would represent a common treatment condition in contrast experiments. A change in sector position would then constitute a manipulation of the separation (and orientation) between test (sector) and inducing (LBG) fields. It would then be expected that test brightness would decay as the separation between test and inducing fields was reduced, with the inhibition of brightness becoming proportionately greater at small separations (Leibowitz, Mote, & Thurlow, 1953). The DBG data of Fig. 1 are consistent with this description.

Unfortunately, there is relatively little data in the contrast literature that is clearly related to the LBG data of Fig. 1. Using a Macbeth disc-annulus type of target configuration, Stevens (1967) demonstrated that increasing the width of the annulus from a small initial value produces a rapid decline in disc brightness that quickly reaches asymptote. A maximal inhibitory effect is exerted upon a 2° 12′ disc by an annulus less than 1° in width. In the present experiment, Sector 6 of the ring, while bordering the DBG, is flanked by a large area of LBG. If the increased embeddedness of Sector 10 within the LBG is analogous to an increase in the area of the inducing field (annulus) in Stevens' experiment, in the vicinity of asymptote, then little change in matched luminance as a function of sector position would be expected. This reasoning produces a conclusion consistent with the LBG divider data of Fig. 1. It also contains the basis for subsequent experiments, i.e., the present experiment, or one like it, should be replicated under conditions permitting the physical isolation of component figure parts. This reasoning, unfortunately, does not provide an explanation for the LBG no-divider data. Perhaps, somewhat paradoxically, it is the existence of a luminance gradient under the no-divider condition that constitutes the demonstration of the good figural properties Koffka claimed for the ring.

REFERENCES

Berman, Phyllis W., & Leibowitz, H. W. Some effects of contour on simultaneous brightness contrast. *Journal of Experimental Psychology,* 1965, **69**, 251–256.

Helson, H. Some factors and implications of color constancy. *Journal of the Optical Society of America,* 1943, **33**, 555–567.

Koffka, K. *Principles of gestalt psychology.* New York: Harcourt, Brace & World, 1935.

Leibowitz, H., Mote, F. A., & Thurlow, W. R. Simultaneous contrast as a function of separation between test and inducing fields. *Journal of Experimental Psychology,* 1953, **46**, 453–456.

Osgood, C. E. *Method and theory in experimenta psychology.* New York: Oxford Universit Press, 1953.

Steger, J. A. Reversal of simultaneous lightness-contrast. *Psychological Bulletin,* 1968, **70**, 774–781.

Stevens, J. C. Brightness inhibition re size of surround. *Perception and Psychophysics,* 1967, **2**, 189–192.

Wallach, H. Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology,* 1948, **38**, 310–324.

(Received February 12, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 410–414

INFORMATION PURCHASE IN A DECISION TASK FOLLOWING THE PRESENTATION OF FREE INFORMATION[1]

GORDON F. PITZ[2] AND HELEN REINHOLD BARRETT

Southern Illinois University

Previous studies have found that *S*s make riskier decisions following the purchase of larger amounts of information. One possible explanation is that the subjective probability of being correct increases with the sample size alone. In three experiments, *S*s were presented with 0, 2, 4, or 6 items of free information and were permitted to purchase additional amounts prior to making a decision. There was statistically significant evidence in at least two of the experiments that less information was purchased following larger amounts of free information, with probabilities held constant, supporting the original hypothesis. However, the magnitude of the effect was slight and was not enough to account for the earlier results.

Many situations requiring a decision involve some degree of uncertainty concerning the eventual outcome, and it is frequently possible to defer the choice until relevant information has been found to reduce the uncertainty. Acquiring relevant information usually entails some cost, monetary or otherwise, however; and it is necessary to balance the increasing cost of additional information against the increasing reduction of uncertainty. This problem has been analyzed within the context of statistical decision theory (e.g., Raiffa, 1968), and a series of papers has investigated the problem from a descriptive point of view (Pitz, 1968, 1969a, 1969b; Pitz, Reinhold, & Geller, 1969).

The typical experimental task requires *S*s to choose between two hypotheses on the basis of relevant but unreliable information. There is a known probability (P) of predicting correctly from any one item of information, and since successive items are conditionally independent of each other, the probability of making a correct decision may be increased by purchasing additional information. A detailed description of the task is given by Pitz et al. (1969). The probability of making a correct decision is a function only of the difference (d) between the number of events predicting each of the two decisions. Hence, the optimal strategy for purchasing information is to request additional items until d reaches some critical value.

One result found in all studies is that *S*s tend to make riskier decisions following the purchase of larger amounts of information than they make after purchasing less information. If a sequence of information is highly diagnostic early, i.e., if most of the early items point to the same decision, *S*s will make a response with a high probability of being correct. If, however, the early items are equivocal, *S*s will settle for making a decision with a smaller probability of being correct. It is as if they use a critical value of d that decreases as the sample size (n) increases. Similar results were observed in a study by Sanders and Ter Linden (1967). The effect of this behavior sometimes is to reduce markedly the amount of money that *S*s can expect to win. This result is in contrast to the effect of merely purchasing a nonoptimal amount of information; purchasing too much or too little information typically has very little effect on the expected monetary gains, providing that a roughly constant critical value of d is used.

Within the normative framework of statistical decision theory, there are at least two possible explanations for the observed decreasing critical d. The Ss may believe that the probability of making a correct decision is greater following larger sample sizes, regardless of the objective probabilities; i.e., the subjective probability for the favored decision may increase as a function of n alone, as well as increase with d. An alternative explanation is that the negative utility for the cost of information increases as a function of sample size, Ss being unwilling to "throw good money after bad"; i.e., Ss may set an upper limit on the amount of information they are willing to purchase if the early information does not lead to a clear decision.

---

[1] This study was supported in part by Grant GS-1497 from the National Science Foundation and in part by a grant from the Graduate School of Southern Illinois University. Experiment OS was part of an MA thesis conducted by the second author under the direction of the first author. Assistance in the other two experiments was provided by E. Scott Geller, to whom the authors wish to express their thanks.

[2] Requests for reprints should be sent to Gordon F. Pitz, Department of Psychology, Southern Illinois University, Carbondale, Illinois 62901.

In order to understand behavior more completely in an information-seeking task, it would be advantageous to know more about Ss' opinions following the purchase of information. In an unpublished experiment,[3] Ss gave confidence judgments for their decisions following the purchase of information. However, the error variance in the results was large, and it was felt that requiring Ss to consider two tasks simultaneously, the purchasing of information and the giving of a confidence judgment, resulted in their paying too little attention to the confidence task. Insofar as a trend was observable in the results, it did not support the hypothesis of a subjective probability increasing as a function of sample size. Confidence in a decision was greater when the objective probability of being correct was greater, but when objective probability was held constant, confidence tended to decrease as a function of sample size. Similar results were observed in an experiment by Pitz (1967), in which the only task was the giving of a confidence judgment. Confidence in a decision increased as a function of d and decreased as a function of n with d held constant.

An alternative way of examining Ss' beliefs that does not require the use of a second task is to provide them with the opportunity of purchasing additional information. It is assumed that the more certain Ss are that a decision would be correct, the less additional information they require before making a decision. In the present experiments, Ss were provided with a variable amount of information free of charge and were then given the opportunity to buy more before making a decision. This procedure permits a distinction to be made between the two hypotheses mentioned earlier. In the earlier studies it was not possible to discriminate between the two, since sample size and investment in money increased simultaneously. However, during the presentation of free information, sample size increases, but the investment in money does not. Optimally, the amount of information purchased following a given amount of free information should be a function only of the objective probabilities (and hence of d) following the free information. However, this would not be the case if subjective probability increases as a function both of d and of n independently of d; the information purchased would vary inversely with both the probability of making a correct decision after the free information and the amount of free information.

METHOD

Procedure.—Details of the procedure may be found in Pitz (1969a, or 1969b). Briefly, Ss were told to predict on each trial which of two hypotheses was correct, each hypothesis having a prior probability of .5. Any number of initial tests could be made before making a decision, and each test would predict the eventual outcome with probability .6. In the scale of values used for this task, tests cost $1 each, correct decisions were worth $100, and incorrect decisions cost $100.

Design.—Information was presented to Ss in two stages. First, a fixed amount of free information was presented, either 0, 2, 4, or 6 items. Each item of free information consisted of a test that was made free of charge; this free, prior information represented the starting point for each decision problem. The amount of free information is referred to as n_p, or prior n; and the difference between the two types following free information is referred to as d_p, or prior d. Following the free information, Ss specified the amount of costly information they desired before making a decision.

Three separate experiments were run, the presentation of information being slightly different in each experiment. The first (Exp. OS) used an "optional stopping" procedure in which Ss purchased the costly information one item at a time, deciding after each item whether or not to request more. This was the procedure used in earlier experiments. The value of n_p was varied randomly from trial to trial. Two experiments (FS_1 and FS_2) used a "fixed stopping" procedure in which, following the free information, *S*s specified the exact amount of costly information they wished to receive; i.e., rather than purchase information one item at a time, the entire sample size was determined in advance. In Exp. FS_1, the amount of free information was varied randomly, as in Exp. OS. In Exp. FS_2, all trials using a given value of n_p were presented in a single block, and the order of blocks was balanced in a Latin square design. Group 1 received free information in the order 0, 2, 4, 6 items; Group 2, the order 2, 6, 0, 4; Group 3, the order 4, 0, 6, 2; and Group 4, the order 6, 4, 2, 0.

---

[3] The experiment was conducted by D. M. Barry. Details of the results are available from the first author.

Since the analysis of data was performed in terms of n_p and d_p jointly, it was necessary that each value of d_p occur reasonably often for each value of n_p. Consequently, more trials were given with the larger values of n_p. In Exp. OS and FS_1, 80 trials were given; n_p of 0, 2, 4, and 6 were presented 8, 16, 24, and 32 times, respectively. In Exp. FS_2, 100 trials were used; and n_p of 0, 2, 4, and 6 were presented 13, 22, 29, and 36 times, respectively.

Forty *S*s were used in Exp. OS and in Exp. FS_1. In Exp. FS_2, four groups of 14 *S*s each were used. All *S*s were male volunteers from an introductory psychology course who received partial course credit for their participation. In addition, they received a small amount of money, ostensively the amount depending upon their performance during the experiment. In fact, to avoid legal problems, all *S*s were later paid $1.70.

RESULTS

The data from Exp. OS were used to determine whether *S*s were indeed giving riskier decisions as sample size increased. The procedure for evaluating strategies of information seeking is described in detail by Pitz et al. (1969), and the detailed analysis of the present data may be found in Reinhold (1968). As in all other studies, it was observed that *S*s behaved as if they adopted a critical *d* that decreased as *n* increased.

The major analysis in all three experiments was a determination of the amount of costly information purchased as a joint function of n_p and d_p. The number of items of costly information, referred to as n_c, should depend for an optimal *S* only on d_p, not on n_p, and should be smaller the larger the value of d_p. Specifically, in fixed stopping experiments, *S*s should have purchased 23 items for $d_p = 0$, 18 for $d_p = 2$, and 9 for $d_p = 4$. In the optional stopping case, n_c should have been a random variable, with an expected value of 19.2 for $d_p = 0$, 15.9 for $d_p = 2$, and 5.8 for $d_p = 4$. Figure 1 shows the average value of n_c as a joint function of n_p and d_p for the three experiments. All showed very similar results; the average amount of information purchased decreased as d_p increased, but was generally much less than was optimal; and with d_p held constant there was a tendency for less information to be purchased following larger values of n_p. There were, however, some reversals of the latter trend.

Since the effect on n_c of n_p with d_p held constant was of major interest, the following procedure was used to test for the significance of the trends shown in Fig. 1. Average values of n_c were computed for each *S* for each n_p and d_p. There are six possible orthogonal comparisons that may be made between means at differing values of n_p and a constant value of d_p. There is one such comparison for d_p of 4, i.e., between n_p of 6 and 4; there are two comparisons possible for d_p of 2, e.g., one for n_p of 6 vs. n_p of 4 and one for the combined n_p of 6 and 4 vs. n_p of 2; finally, there are three possible comparisons at d_p of 0. Defining a "success" as an orthogonal comparison for which the average n_c was larger for the smaller n_p, a count was made of the number of "successes" (0–6) for each *S*. The frequency of *S*s in each of the seven categories was tabulated, and a chi-square test was made of the null hypothesis that the frequencies were binomially distributed with probability .5. The procedure may be clarified by reference to Table 1, in which the relevant frequencies are tabulated. In order to obtain the minimum required expected frequencies in each category, it was necessary to combine *S*s with 0 and 1 "successes" and with 5 and 6 "successes." Since there were, then, five categories, the chi-square statistic had 4 *df*. The data from the four groups of Exp. FS_2 were combined for this analysis since the results for the four groups did not appear to differ.

Using this procedure, the smaller mean

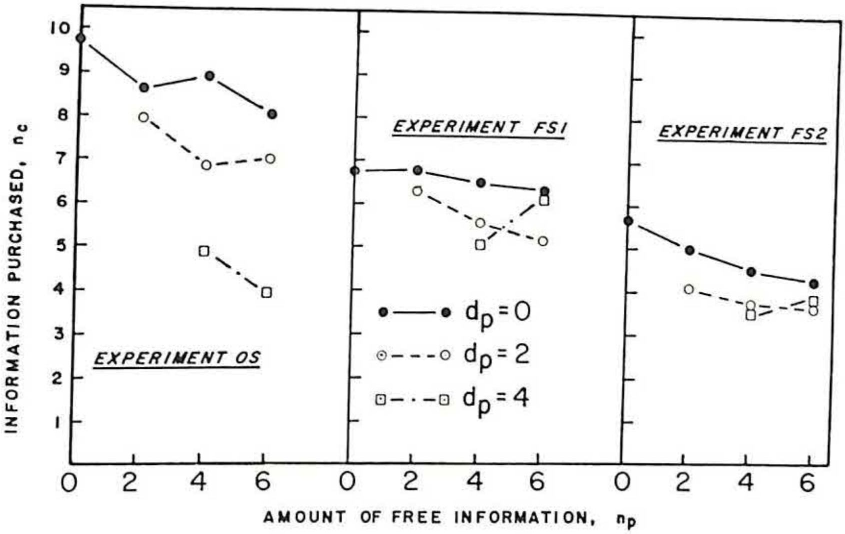


Fig. 1. Mean amount of costly information (n_c) purchased as a function of n_p and d_p in the three experiments.

n_c for larger n_p was significant for Exp. OS and FS_1, χ^2 (4) = 31.27 and 28.45, respectively, $p < .01$. For Exp. FS_2, χ^2 (4) = 12.04, $p < .05$, and for the four separate groups of Exp. FS_2, the individual chi-square values were not significant at any reasonable level of p.

DISCUSSION

The results of at least two of the three experiments provide statistically significant evidence that the amount of information purchased decreased as the amount of free information increased, with objective probabilities held constant. In absolute terms, however, this effect was slight; and with the large number of *S*s used in the experiments, the power of the statistical tests was high. The conclusion must be that if subjective probability does indeed increase solely as a function of sample size, the effect must be small. The increase in subjective probability implied by the results of Fig. 1 is not nearly sufficient to account for the degree of increase in risk of decisions observed in other studies. One may conclude, then, that in optional stopping tasks, *S*s make riskier decisions as n increases largely because of limits on the amount of information they will purchase. One problem remains, however. The trend of the results as a function of n_p in Fig. 1 is the reverse of that found when *S*s gave a judgment of confidence in a decision (Pitz, 1967). That is, *S*s' confidence decreased as a function of n, and yet they required less additional information as a function of n in the present study. The reason for this apparent contradiction is not clear.

TABLE 1

Frequency of *S*s in Each Experiment with Each Number of "Successes"

Number of "successes"	Exp.		
	OS	FS_1	FS_2
0 or 1	0	2	6
2	6	5	6
3	8	10	14
4	11	9	18
5 or 6	15	14	12

Note.—"Success" is defined as one orthogonal comparison for which n_c was larger for the smaller n_p.

The *S* who is behaving optimally would allow the purchase of costly information to depend only on the objective probabilities following the free information. In the present experiments, the objective probability is related directly to d_p, the difference between the two kinds of events following free information; the larger the value of d_p, the less subsequent information should be requested. It may be observed from Fig. 1 that *S*s did indeed respond in this way, and the effect was present for all *S*s almost without exception. However, under most conditions there was a gross underbuying of information. For $d_p = 0$, *S*s reduced their expected gains by more than one-third as a result of this underbuying, while as d_p increased, behavior became more optimal. Looking at it another way, the extent to which n_c decreased as a function of d_p was considerably less than optimal. The effect of d_p on information purchase was much greater under conditions of optional stopping, when *S*s purchased information one item at a time, than it was under conditions of fixed stopping, when they specified beforehand the amount of information they wished to purchase. It is likely that the fixed stopping procedure encouraged *S*s to think in terms of constant sample sizes. It was not the intent in the present study to compare OS and FS procedures; such a comparison has been made by Fried and Peterson (1969). It might be noted, however, that performance was generally more optimal in the OS condition.

REFERENCES

Fried, L. S., & Peterson, C. R. Information seeking: Optional versus fixed stopping. *Journal of Experimental Psychology,* 1969, **80**, 525–529.

Pitz, G. F. Sample size, likelihood, and confidence in a decision. *Psychonomic Science,* 1967, **8**, 257–258.

Pitz, G. F. Information seeking when available information is limited. *Journal of Experimental Psychology,* 1968, **76**, 25–34.

Pitz, G. F. The influence of prior probabilities upon information seeking and decision making. *Organizational Behavior and Human Performance,* 1969, **4**, in press. (a)

Pitz, G. F. Use of response times to evaluate strategies of information seeking. *Journal of Experimental Psychology,* 1969, **80**, 553–557. (b)

Pitz, G. F., Reinhold, H., & Geller, E. S. Strategies of information seeking in deferred decision making. *Organizational Behavior and Human Performance,* 1969, **4**, 1–19.

Raiffa, H. *Decision analysis: Introductory lectures on choices under uncertainty.* Reading, Mass.: Addison-Wesley, 1968.

Reinhold, H. Information purchase following the presentation of free information. Unpublished masters thesis, Southern Illinois University, 1968.

Sanders, A. F., & Ter Linden, W. Decision making during paced arrival of probabilistic information. *Acta Psychologica,* 1967, **27**, 170–177.

(Received February 14, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 415–420

TRANSFORMATIONS ON SETS IN SHORT-TERM MEMORY SET SIZE REDUCTION BY DELETION[1]

DONALD V. DEROSA[2]

Bowling Green State University

An individual's ability to delete items from a memorized set was examined in a short-term recognition memory task. A first (positive) set of four different items was followed by a second (deletion) set. Items in the second set that were identical to items in the positive set were to be deleted from the positive set. A test stimulus was presented following the last item to be deleted, and *S* made a positive response if the test element was identical to any item in the set of items that remained positive and a negative response otherwise. Reaction time (RT) to the test item indicated: (*a*) that *S*s performed a deletion operation and (*b*) that the deletion operation was most difficult when the deletion set was composed of both items to be deleted and items not present in the positive set. Attention was given to an examination of the results in terms of strategies of memory search.

Posner (1967) has emphasized the importance of examining the functions that short-term memory (STM) plays in human information processing. One use of an STM mechanism he describes is that of operational memory, where information in long-term memory is activated and used in problem solving. This use of STM is similar to the "workspace" notion described by Hilgard and Bower (1966) and the active memory proposed by Sternberg (1968).

If STM is used as a workspace where mental operations are performed, it may be possible to examine an individual's ability to perform a transformation on a set of items that has been recently perceived. In the present investigation, the principal operation investigated was deletion.

The task is a variation of a recognition memory task used by Sternberg (1966). In his investigation, a trial consisted of a serial presentation of a subset of digits drawn from the set of 10 single digits (0–9). The presented digits defined a set referred to as the *positive* set, while its complement (digits not presented) was referred to as the *negative* set. Following the last item of the positive set, a probe digit was shown, and *S* was to respond with a first (positive) response if the probe stimulus was identical to any member of the positive set and with a second (negative) response otherwise. Sternberg varied the size of the positive set and examined reaction time (RT) to positive and negative probes as a function of the size of the positive set. Reactions to both positive and negative probes increased linearly as a function of the size of the positive set. Moreover, the slopes of both functions were the same.

Sternberg (1966) views these results as support for a serial exhaustive model of memory search. The model proposes that an internal representation of the test stimulus is compared successively to symbols in memory, each comparison resulting in either a match or a mismatch. The time from the beginning of one comparison to the beginning of the next (the comparison time) has the same value for successive comparisons. A positive response is made if there has been a match and a negative response otherwise. The search is serial since comparisons are made successively. The search is exhaustive because the search *does not* terminate when a match has been made, but continues until all possible comparisons have been made. Therefore, linear functions relating search time to size of the positive set should yield equal slopes for both positive and negative responses.

---

[1] This paper is based on a thesis submitted to Kent State University in partial fulfillment of the requirements for a PhD degree. The study was conducted while the author was a United States Public Health Service predoctoral fellow. The author wishes to express appreciation to his sponsor, Robert E. Morin. Also, the helpful suggestions of Bert Forrin are acknowledged. The research was supported in part by National Science Foundation Grant GB 4343 awarded to Robert E. Morin.

[2] Requests for reprints should be sent to Donald V. DeRosa, Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43402.

The subtraction task.—In the subtraction task, the enumeration of the positive (P) set on one display cell is followed by a second series of digits presented on a second display cell. The second sequence defines the deletion (D) set. Those items in D that are identical to members of P are to be deleted from P and regarded as negative-set items. Those items in D which are not identical to P elements are also regarded as negative items.

Table 1 provides an illustration of a single trial. The digits of P are shown serially and followed by a sequential presentation of D items. After the termination of D, a single test stimulus or probe is displayed. In the example in Table 1, the P set is reduced from an original set (4, 7, 9, 3) of Size 4, $n(\mathrm{P}) = 4$, to a net P set (4, 9), of Size 2, $n(\mathrm{P} - \mathrm{D}) = 2$. The D set (7, 3, 5) contains three digits, $n(\mathrm{D}) = 3$, two digits (7, 3) which are in P and a third (5) which is not in P.

TABLE 1

ILLUSTRATION OF A SINGLE TRIAL

Serial position							Some possible probes	
P Set				D Set				
1	2	3	4	1	2	3	Probe	Correct response
4	7	9	3	7	3	5	4 8 5 7	Yes No No No

Note.—P = positive; D = deletion.

The digits in the D set can be described as active ($\mathrm{D_A}$) and inactive ($\mathrm{D_I}$) items. Presentation of an item in P "activates" an item; on subsequent presentation of the same item in D, it is an *active* item that is being deleted from P. Consequently, items in D that also appeared in P are termed $\mathrm{D_{Active}}$ or $\mathrm{D_A}$ items. An item not presented in P is *inactive* when it is shown in D and is therefore a $\mathrm{D_{Inactive}}$ or $\mathrm{D_I}$ item. In Table 1, 7 and 3 are active items, since they were presented in P, whereas 5 is inactive. Probes may be selected from among four types of items based on a digit's state prior to the D set and its state following the D set. The four types of items and the reasons for their being positive or negative are as follows: (*a*) An element will be positive if it was a member of P but not D ($\mathrm{P\overline{D}}$); (*b*) an element will be negative if it was not a member of P or D ($\mathrm{\overline{P}\overline{D}}$), a member of D but not P ($\mathrm{\overline{P}D}$), or both a member of P and a member of D (PD). Probe Stimulus 4 illustrates the only condition under which an item will be positive at the time of the test. Probes 8, 5, and 7 are negative probes, which illustrate in the preceding order the circumstances under which a negative response is appropriate.

It is noteworthy that correct responding is not contingent on the performance of a deletion operation. A nondeletion strategy might involve the storage of all presented items, with each item being tagged as a P or D element. When a probe is presented, a search of a memorial representation of the digits and their tags could be initiated to determine the present state of the test digit.

It is possible to formulate alternative search strategies that might accompany a nondeletion operation. For example, *S*s might first compare the test digit to each P item (forward exhaustive search) at the end of the search, and a "no" response would be initiated if a match has not been detected. If a match is made, a subsequent search of D is necessary to determine if the item is also present in D. If the test item is in D, then a "no" response is appropriate; otherwise, "yes" is the correct response. Assuming that each comparison takes a constant unit of time and that the search through the subsets is exhaustive, reactions to elements which were members of P ($\mathrm{P\overline{D}}$ and PD) should be slower than reactions to items not present in P ($\mathrm{\overline{P}\overline{D}}$ and $\mathrm{\overline{P}D}$) since under this model a search of D is necessary following a match in P. Since an exhaustive search of D is also assumed, no difference in RT should exist for the $\mathrm{P\overline{D}}$ and PD probe types. Each would result in the maximum number of comparisons: $n(\text{P}) + n(\text{D})$.

The model described here is illustrative. It defines a nondeletion strategy for a forward (P searched first) exhaustive search. Other nondeletion strategies lead to different predictions for the various probe types. If *S*s use a nondeletion strategy, the determination of the characteristics of the search employed should be reasonably straightforward.

All nondeletion strategies place an increasing load on *S* as the size of D increases. When $n(\text{P}) = 4$ and $n(\text{D}) = 1$, 2, or 3, *S* must store a total of 5, 6, and 7 P and D items, respectively. The *S*, therefore, in no way takes advantage of the possible reduction in memory load by eliminating P items that are presented in D. Nonreduction strategies therefore would not predict a decrease in RT to positive probes as a function of decreases in $n(\text{P} - \text{D})$ produced by the presence of active items in D. A deletion strategy, on the other hand, might allow *S* to compare the probe digit to a P set reduced in size by the previous deletion of items from the P set. Reaction time should be sensitive to this reduction if *S*s compare the probe digit to a smaller P − D set.

In the present study, one question of major concern is whether the reduction in Size P by deletion will be reflected in faster RTs for smaller P − D sets. It was noted previously that the RTs for positive and negative reactions had been shown to be a positive linear function of size of the positive set (Sternberg, 1966). In this study, positive and negative reactions can be examined as a function of $n(\text{P} - \text{D})$. If *S*s search a reduced positive set in the same way that they would search a P set not followed by any D set, similar slopes should be evidenced for functions relating positive-set size to RT for reduced and unreduced P sets of equal net size.

METHOD

Subjects.—The *S*s were 16 male undergraduate volunteers, who were paid for their participation in the experiment. All *S*s served in each condition of the experiment.

Apparatus.—The *S*s were tested individually in a sound-deadened isolation chamber. Stimuli were presented on the viewing faces of three rear-projection display cells (I. E. E. Series 10,000), which were mounted at eye level on a panel approximately 3 ft. in front of *S*. Two cells were horizontally adjacent. The left cell provided the presentation of the *P* set. The D set was shown on the right cell. A probe cell was mounted below and centered between the two upper cells.

A microphone monitored *S*s' vocal responses and also served to activate a voice key. The *S*s were instructed to say "yes" if the probe digit was identical to any item in the P − D set and "no" otherwise. The voice key turned off the stimulus, advanced the paper tape reader, and activated a digital recorder, which printed RTs to the nearest .001 sec.

Experimental conditions and procedure.—The experimental conditions are summarized in Table 2.

TABLE 2

SCHEMATIC REPRESENTATION OF EXPERIMENTAL CONDITIONS

Cond.	n(P)	n(D)	n(P − D)
1-0-1	1	0	1
2-0-2	2	0	2
3-0-3	3	0	3
4-0-4	4	0	4
4-1-3	4	1	3
4-1-4	4	1	4
4-2-2	4	2	2
4-2-3	4	2	3
4-2-4	4	2	4
4-3-1	4	3	1
4-3-2	4	3	2
4-3-3	4	3	3
4-3-4	4	3	4

Each condition is described by three numerals. The first numeral corresponds to the size of P, $n(\text{P})$. The second designates the size of D, $n(\text{D})$. The net P set size is indicated by the third numeral, $n(\text{P} - \text{D})$. For example, in Cond. 4-1-3, $n(\text{P}) = 4$, $n(\text{D}) = 1$, and $n(\text{P} - \text{D}) = 3$. In this condition, the single digit in D had to be an active (D_A) item since P is reduced to three items.

Stimuli from the P and D sets were white on a black ground and were presented at a rate of 3/sec Each stimulus was displayed for 200 msec. and was followed by a 130-msec. off time. There was no delay between the last digit of P and the first item in D. The time between the termination of the last digit of D and the onset of the probe was 750 msec. The probe stimulus remained on until *S* responded. A 1.5-sec. intertrial interval occurred between *S*'s response and the beginning of the next trial.

Ten program tapes were prepared to present blocks of trials to *S*s. Within each tape or list, $n(\text{D})$ was constant at 0, 1, 2, or 3. When D contained D_A items, these were selected systematically such that items at each serial position in P were represented equally often in D. Inactive (D_I) items in D were chosen randomly from among those digits not present in P. The serial positions within D of the selected D_A and D_I items were varied systematically, so that each item appeared equally often at each serial position in D.

Two constraints were imposed on the choice of a probe stimulus. First, equal probabilities of positive and negative probes were maintained by equating the number of probes identical to items in P − D with the number of probes for items not in P − D. Secondly, probe stimuli were identical to items in each serial position of P and D equally often.

With these constraints, 400 trials were selected for use in the experiment. The 400 trials were divided into 10 lists or blocks of 40 trials in the following manner: A first list randomly intermixed trials for Cond. 1-0-1, 2-0-2, 3-0-3, and 4-0-4, the base-line conditions for which there were no deletion items. A second list presented trials in which $n(P) = 4$ and $n(D) = 1$. Four lists were prepared for conditions in which $n(P) = 4$ and $n(D) = 2$, and 4 were prepared for conditions when $n(D) = 3$. Within each list, trials representative of the different conditions were randomly arranged.

Each block of trials was preceded by 4 lead-in trials, which had the same characteristics as the 40 experimental trials. Data from the lead-in trials were recorded but not included in the data analyses.

Pretraining and pretesting session.—One week before the experiment, Ss reported for a 35-min. pretraining and pretesting session. During this session, Ss were acquainted with the manner in which P and D items would be displayed. Each probe remained on until *S* responded. Instructions encouraged *S*s to respond rapidly and accurately. The *S*s were given four blocks of 20 trials each, one block corresponding to each D-set size. In the first block, $n(P)$ was either 1, 2, 3, or 4 and $n(D)$ was 0. In the three remaining blocks, $n(P)$ was always 4 and $n(D)$ was either 1, 2, or 3. The four blocks were first presented at a rate of two digits per second in the order $n(D) = 0$, 1, 2, and 3. Following the fourth block, the stimuli were presented again in the same order at a rate of three digits per second, a rate identical to that of the actual experiment.

For each *S* in each block, mean RTs for correct responses and error percentage were tabulated. Of the 19 *S*s who participated in pretraining, 3 were not invited to participate in the experiment. Their error rates were 35%, 47%, and 53% when $n(P) = 4$ and $n(D) = 3$ at the fast presentation rate. All other *S*s had error rates of less than 25% in this condition. The mean error rate over all trials of pretraining was 12% for all 16 *S*s.

Experimental sessions.—The 10 experimental blocks were presented once in a 1-hr. experimental session. Four sessions were run for each *S* on 4 consecutive days. The order of presentation for the 10 blocks was determined randomly each day for each *S*. Each session began with a warm-up block of 20 trials. This block was identical to the pretraining block, where $n(P) = 4$ and $n(D) = 1$. These trials were scored for accuracy and speed, but not included in the data analysis.

Prior to a block of trials, *S*s were always told the number of items that would appear on the left cell (P) and the number of items that would appear on the right cell (D). For the block with $n(D) = 0$, *S*s were informed that P would be either one, two, three, or four digits followed by a probe stimulus. Each block of trials was preceded by a green warning signal shown on the upper left cell (P cell). The warning signal remained on until *S* said, "Ready," and then the trials began. Each block lasted approximately 3.5 min. A 2-min. rest period followed each block, and a 5-min. rest was given after the fifth block. Over the four experimental sessions, 40 blocks of trials were presented and 1,600 responses (40×40) were obtained from each *S*.

RESULTS

The principal analyses of the data were based on mean RTs for correct responses computed separately for each *S*. None of the *S*s had error rates higher than 25%, and the average error rate for the entire experiment was approximately 10%.

Positive sets not followed by deletion sets.—The RT results for the base-line conditions are shown in Fig. 1. Here, RT to positive and negative probes is plotted as a function of $n(P)$ when P varied and no D items were shown. The most noteworthy feature of these data is the linear relationship between $n(P)$ and RT. All 16 *S*s evidenced individual curves like the average curves in that all had a positive slope. The $n(P)$ variable was significant, $F(3, 45) = 115.38$, $p < .001$. In addition, only the linear component of $n(P)$ reached significance, $F(1, 15) = 148.59$, $p < .001$. All orthogonal components of the $n(P) \times$ Responses interaction failed to reach significance, indicating that the two curves in Fig. 1 do not differ significantly in any aspect of trend. The slope for positive reactions is 45 msec. per unit increase in $n(P)$, and the comparable function for negative reactions is 46 msec. This result is consistent with Sternberg's (1966) serial exhaustive model of memory search.

Deletion sets containing all active or all inactive items.—The left panel of Fig. 2 depicts the effect of size of D, $n(D)$, when all items in D are active. As $n(D)$ varies from 3 to 1, the net number of items that remain positive varies from 1 to 3; larger D sets result in smaller $n(P - D)$ sets when all D items are active.

Each curve in the left panel of Fig. 2 corresponds to a unique type of probe (either $\mathrm{P}\overline{\mathrm{D}}$, PD, or $\overline{\mathrm{PD}}$). The $\mathrm{P}\overline{\mathrm{D}}$ curve represents means for those trials when the probe digit was identical to an item in P − D. The $\overline{\mathrm{PD}}$ function represents the means for those trials when the probe digit was not identical to any digit in P or in D. Both the $\mathrm{P}\overline{\mathrm{D}}$ and the $\overline{\mathrm{PD}}$ curves demonstrate the typical relationship between RT and size of the positive set: as $n(\mathrm{P} - \mathrm{D})$ increases, RT increases. The slopes of these two functions are 47.0 and 34.0 msec. per unit increase in $n(\mathrm{P} - \mathrm{D})$ for the $\mathrm{P}\overline{\mathrm{D}}$ and $\overline{\mathrm{PD}}$ probes, respectively.

The curve labeled PD represents negative reactions to probes present in both P and D. This curve departs noticeably from the relationship between RT and $n(\mathrm{P} - \mathrm{D})$ observed for the other curves in the figure. Specifically, the PD curve initially increases as $n(\mathrm{P} - \mathrm{D})$ increases from 1 to 2, but then drops sharply from 2 to 3.

The variance attributable to $n(\mathrm{P} - \mathrm{D})$ was divided into linear and quadratic components, each of which was significant, linear $F\ (1,\ 15) = 14.07$, $p < .01$; and quadratic $F\ (1,\ 15) = 14.89$, $p < .01$. A separate analysis including only $\mathrm{P}\overline{\mathrm{D}}$ and $\overline{\mathrm{PD}}$ yielded no $n(\mathrm{P} - \mathrm{D}) \times$ Type of Probe interaction, $F\ (2,\ 30) = 1.96$, $p > .05$. It is noteworthy that these two probe types are most like the positive and negative probes employed when a positive set was not followed by D items (see Fig. 1). The

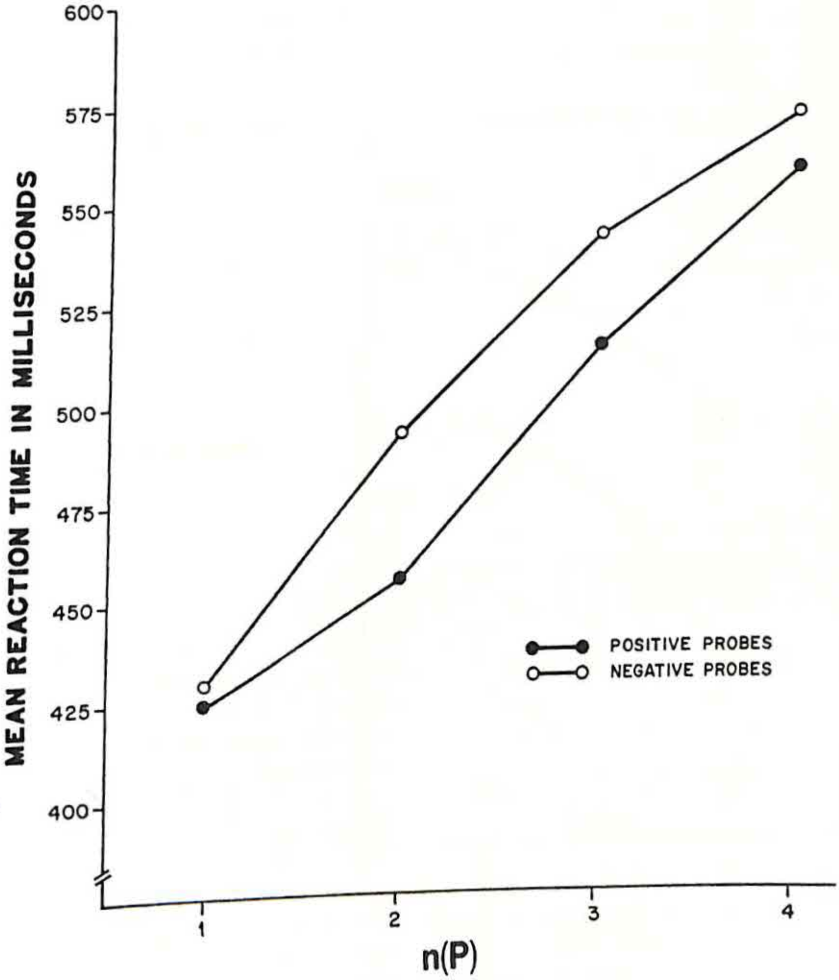


FIG. 1. Mean RT to positive and negative probes as a function of the size of the positive set, $n(\mathrm{P})$, when P was not followed by D items.

$P\bar{D}$ curve represents reactions to items that were positive and presented once. The $\overline{PD}$ curve represents negative reactions to items that were never seen in P or D. The PD curve, on the other hand, represents reactions to items that were formerly positive, but whose state was changed due to presence in D.

Apart from the PD curve, the data reflect the importance of size of the net positive set on RT. Furthermore, the data are consistent with a hypothesis that *S* is searching a reduced positive set exhaustively since the curves for $P\bar{D}$ and $\overline{PD}$ are parallel.

Finally, the absolute level of curves in the left portion of Fig. 2 is about 100 msec. higher than the level for the two curves in Fig. 1. This increase in vertical displacement for positive sets reduced in size by D sets may be a consequence of the additional time in store for remaining P items since these were followed by D items. Items residing in STM for longer periods of time may be less accessible than recently perceived items.

The data for reactions to positive and negative probes when all items in D are inactive are shown in the right panel of Fig. 2. In this figure, $n(\mathrm{D})$ varies, but $n(\mathrm{P} - \mathrm{D})$ is constant at 4 since none of the items in P appear in D. Reactions to $P\bar{D}$, $\overline{PD}$, and PD probes are illustrated in the figure. A $\bar{P}D$ probe is one which is identical to a member of D, but not a member of P.

A striking feature of the data in the right panel of Fig. 2 is the absence of a significant slope in the three curves. All three functions are essentially flat, with slopes of 8, 9, and 7 msec. per unit increase in $n(\mathrm{D})$ for $P\bar{D}$, $\overline{PD}$, and $\bar{P}D$, respectively. The results of an analysis performed on these data in-

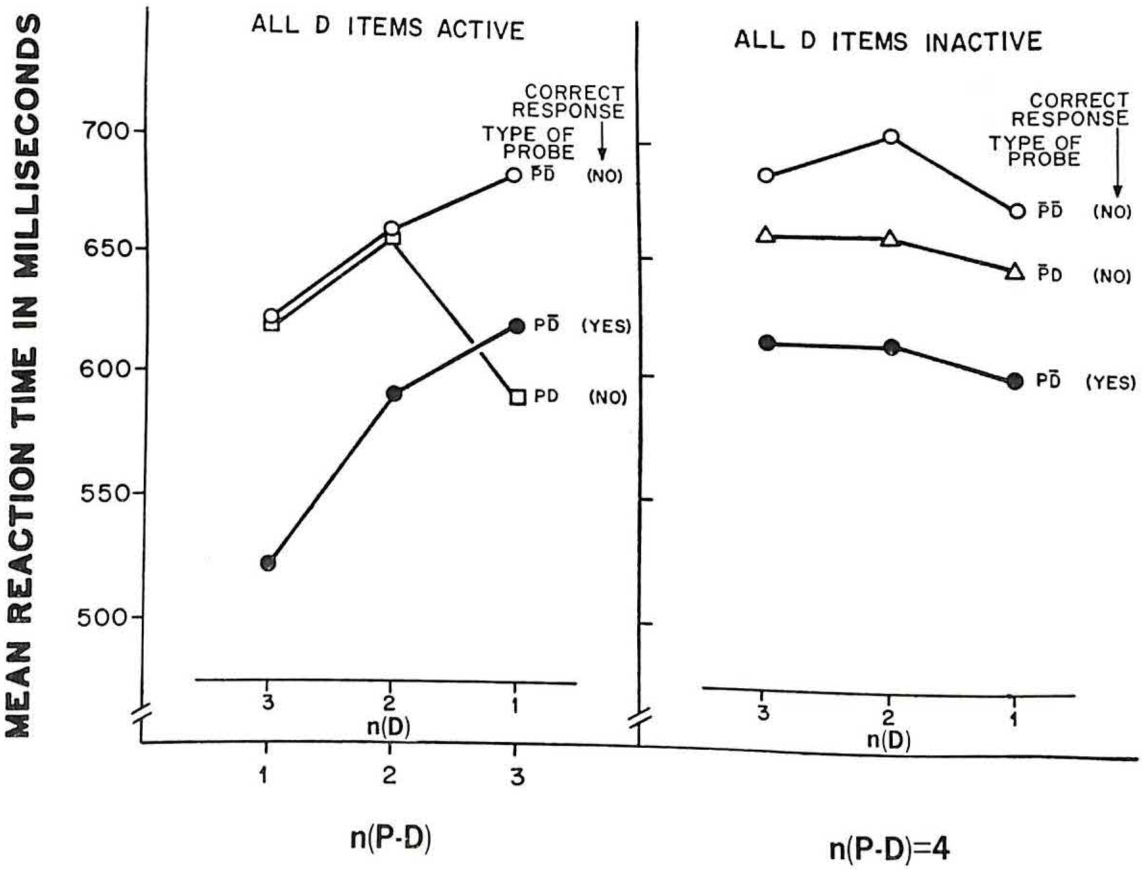


FIG. 2. Mean RT to positive and negative probes as a function of the $n(\mathrm{D})$ when all D items are active (left panel) or all D items are inactive (right panel).

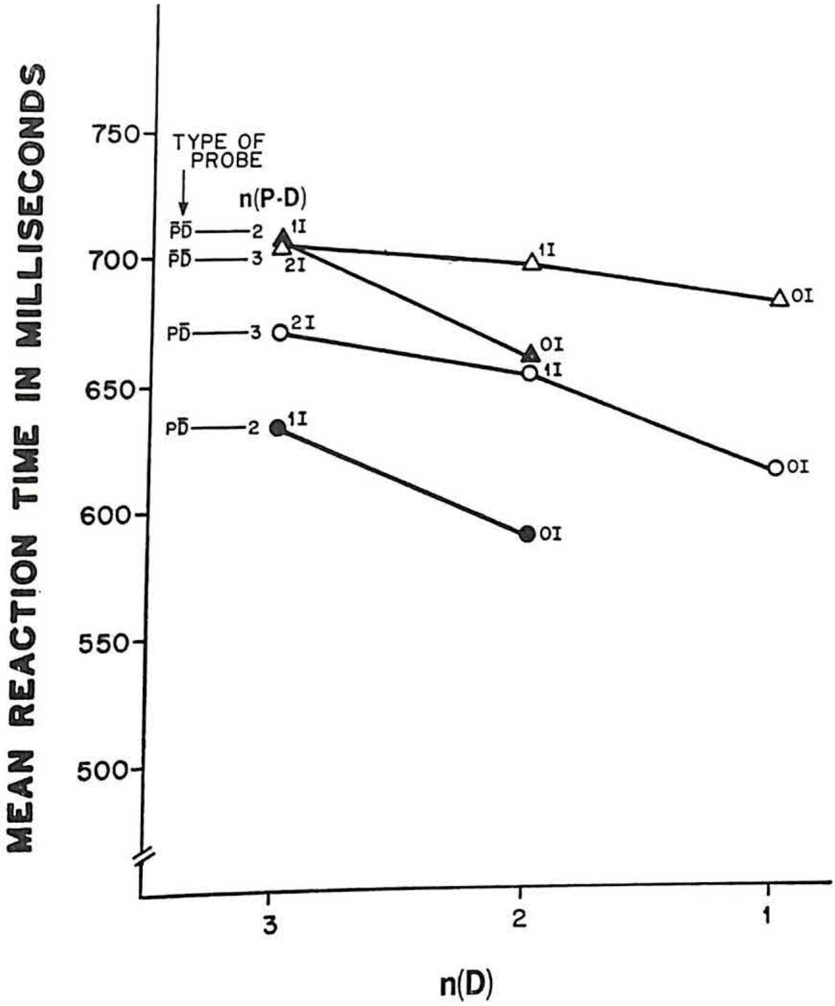


FIG. 3. Mean RT to positive ($P\bar{D}$) and negative ($\overline{PD}$) probes as a function of n(D) when D sets are composed of both active and inactive items. (The number of inactive items present in D is given next to each data point.)

dicated that the only significant variable was type of probe, F (2, 30) = 12.609, $p < .001$. A separate analysis including only the two negative probe types ($\overline{PD}$ and $\bar{P}D$) indicated that the two types of probe did not differ significantly, F (1, 15) = 3.42, $p > .05$. All other main effects and interactions were not significant. These data indicate that when $n(P - D)$ is constant and n(D) varies, RT is not influenced by the number of elements in D.

Deletion sets containing both active and inactive items—Figure 3, like Fig. 2, shows RT as a function of the n(D). In this instance, the data summarize positive and negative reactions to items which did not appear in D. For the first time, data are plotted for conditions in which both active and inactive items were included in the D set.

Interpretation of the curves in Fig. 3 is facilitated by an example. One may consider a positive set, 8-2-3-7. If followed by a D set, 2-7, reactions to a positive probe of 3 would be included in the data for the lower point on the lowest curve since $n(D) = 2$, $n(P - D) = 2$, and there are zero inactive items in D (see legend beside data point). If the D set is changed to 2-5-7, the addition of one inactive item (5) creates a condition in which $n(D) = 3$ and $n(P - D) = 2$ (upper data point on lower curve in Fig. 3). In other words, the points on any one curve in Fig. 3 show the effects of interpolating increasing numbers of inactive terms between P and the probe. These effects are shown separately for positive responses ($P\bar{D}$) and negative responses ($\overline{PD}$) and conditions in which there are *two* D_A items, $n(P - D) = 2$, and *one* D_A item, $n(P - D) = 3$, in the D set.

The general finding is clear: Increasing the size of D by increasing the number of inactive items in D lengthens RT. All curves in Fig. 3 have a positive slope. For $P\bar{D}$ and $\overline{PD}$, the slopes are 44.0 and 46 msec., respectively, when $n(P - D) = 2$. When $n(P - D) = 3$, the comparable figures are 26.0 and 10.0, respectively. An analysis performed on the data for $n(D) = 2$ and 3 repeats the earlier finding that $n(P - D)$ has a significant effect on RT, $F(1, 15) = 10.09$, $p < .01$.

The effect of D sets containing both active and inactive items is depicted again in Fig. 4, but in a different manner. In this figure, positive reactions are shown for all combinations of P and D as a function of $n(P - D)$. The three upper curves are for $n(P) = 4$ followed by either 1, 2, or 3 D items. The number of D_I items present in D is shown beside each data point for the three functions. The lower curve is identical to the lower curve in Fig. 1 and shows positive reactions when $n(P)$ was either 1, 2, 3, or 4 items but was not followed by a D set. The three upper curves were each analyzed separately. The curve for $n(P)$

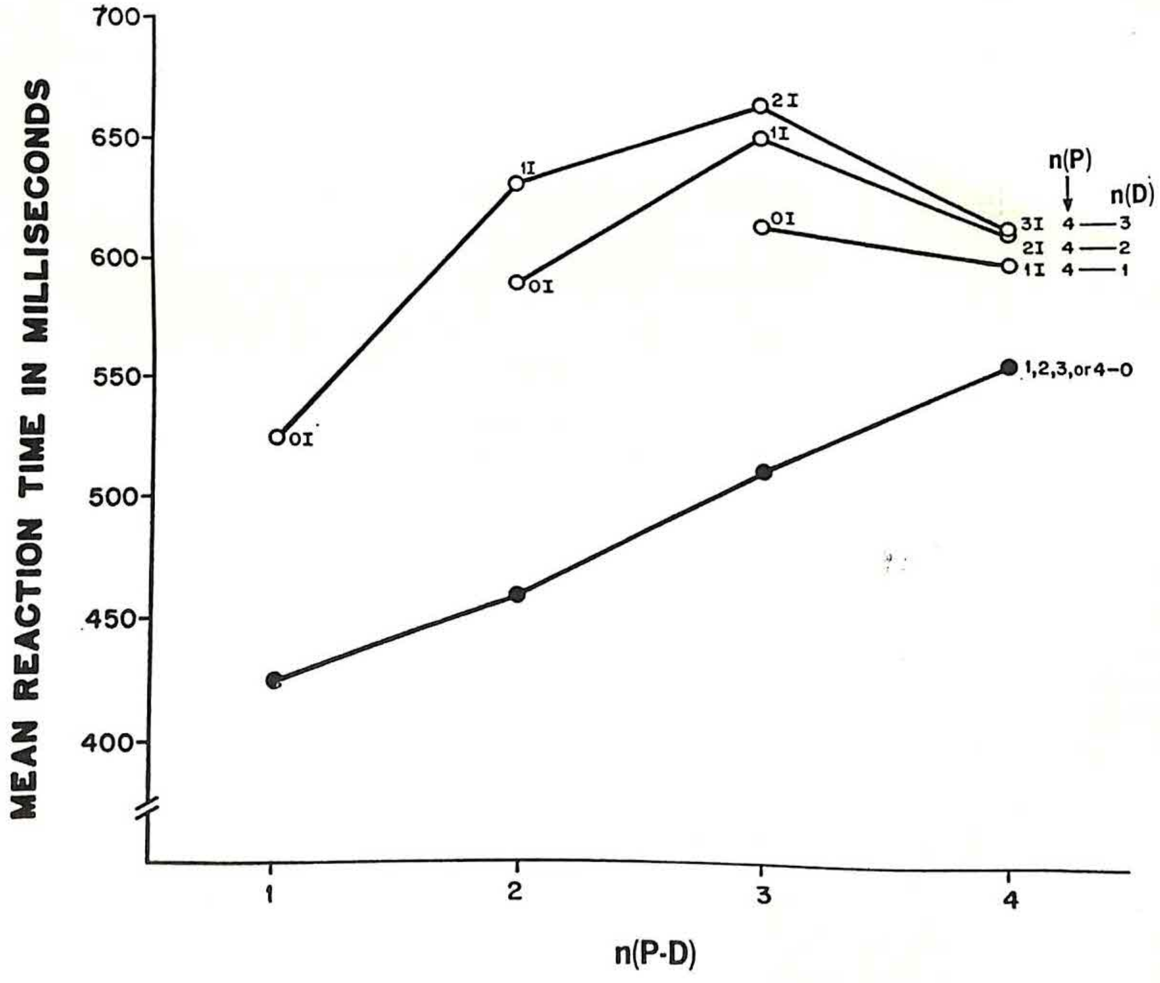


FIG. 4. Mean RT to positive probes as a function of $n(P - D)$. (The number of inactive items present in D is given next to each data point.)

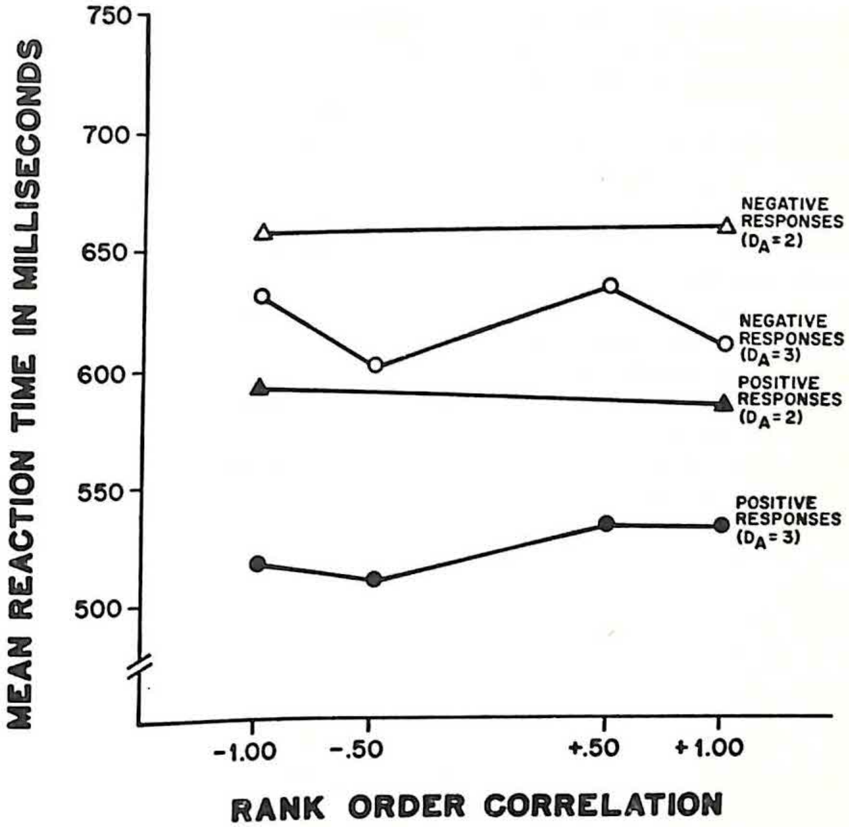


FIG. 5. Mean RT to positive and negative probes as a function of the correlation between the ordinal positions of identical items in P and D.

$= 4$ and $n(D) = 1$ does not have a significant slope, $F\ (1, 15) = .878$, $p > .05$. Similarly, the function for $n(P) = 4$ and $n(D) = 2$ does not have a significant linear component, $F\ (1, 15) = 1.94$, $p > .05$, but it does have a reliable quadratic component, $F\ (1, 15) = 33.99$, $p < .001$. Finally, the uppermost function for $n(P) = 4$ and $n(D) = 3$ has both significant linear, $F\ (1, 15) = 15.25$, $p < .01$, and quadratic components, $F\ (1, 15) = 44.41$, $p < .001$.

These results demonstrate two points. First, they again point out that positive-set size is an important determinant of RT. Secondly, the quadratic components evidenced in the 4-2 and 4-3 curves illustrate the importance of the presence of both D_A and D_I items in D. It is especially clear from the uppermost curve that processing two types of D items on the same trial is more difficult for *S*s than processing sets in which the D items are homogeneous, either all active or all inactive.

Order correspondence between P and D items.—The serial nature of the deletion task encourages an examination of the order correspondence between identical items in P and D. Perhaps deletion would be a simpler task if the order of items in D is identical to the order in which items to be deleted were shown in P.

The effects of the order correlation between items in P and identical items presented in D are shown in Fig. 5. Rank correlations were computed to describe the consistency of the order in which active items occurred in D with the order in which the same items appeared in P. For example, if P were 4, 7, 9, 3 and D were 4, 9, 3, the value of the rank order correlation between the two sets would be + 1.00. The 7 present in P but not D is disregarded in calculating the correlation coefficient. For P sets of Size 4 followed by a D set containing three active items, four possible coefficients exist, + 1.00, + .50, − .50, and − 1.00.

When P sets of Size 4 precede a D set having only two active items and no inactive items, the correlation coefficient is either + 1.00 or − 1.00.

The data summarized are for D sets containing only active items. Four curves are depicted. The two curves with triangles as data points represent reactions when $n(\mathrm{P}) = 4$ and D was two $\mathrm{D_A}$ items. The two remaining curves, with circles as data points, represent reactions when $n(\mathrm{P}) = 4$ and D contained three active items.

The correlation variable appears to have no influence on RT. All four curves in Fig. 5 are essentially flat. The results of two analyses performed separately on the data for P sets followed by two active items and P sets followed by three active items indicated no significant effect of the order variable: F (1, 15) = .84, $p > .05$, for $\mathrm{D_A} = 2$; and F (3, 45) = 2.27, $p > .05$, for $\mathrm{D_A} = 3$.

TABLE 3

SUMMARY OF ERROR DATA

Cond.	n(P)	n(D)	n(P − D)	Mean percentage of errors	
				Positive probes	Negative probes
1	1	0	1	1.	1.6
2	2	0	2	2.0	1.2
3	3	0	3	3.9	1.6
4	4	0	4	5.8	1.4
5	4	1	3	10.2	6.8
6	4	1	4	8.4	5.2
7	4	2	2	13.5	12.1
8	4	2	3	15.8	9.9
9	4	2	4	7.9	6.6
10	4	3	1	14.4	9.7
11	4	3	2	20.3	11.3
12	4	3	3	17.8	9.1
13	4	3	4	12.0	5.4

Summary of errors.—Table 3 summarizes the errors for each condition of the experiment. The mean percentage of errors is shown for positive and negative probes. The errors for positive probes reflect the failure to recognize a probe as identical to a positive item. The errors for negative probes constitute the false detection of a positive item.

In general, the pattern of errors parallels the RT data. Those conditions yielding the longest RTs are typically the same as those that evidence the largest proportion of errors. For example, it was noted that positive RTs were longest when D contained both active and inactive items. In Table 3, Cond. 8, 11, and 12 correspond to conditions in which both active and inactive items appeared in D. Errors are generally most frequent in these conditions.

A second characteristic of the data is the fact that errors are more frequent for positive probes than for negative probes in all 13 conditions. The more frequent error (saying "no" to a positive probe) represents an error of omission. The *S* mistakenly fails to find an item in the P set that is actually there. The errors to negative probes constitute errors of commission. The *S* finds something in P that is not there.

DISCUSSION

The results demonstrate that when P sets are not followed by D items, reactions to both positive and negative probes increase linearly as a function of $n(\mathrm{P})$ (see Fig. 1). Furthermore, the equality of the slopes for positive and negative reactions is consistent with the serial exhaustive search model proposed by Sternberg (1966).

When P sets are followed by D sets containing all active or all inactive elements, the data are generally consistent with the proposition that *S*s do perform a deletion operation. In support of this hypothesis is the observation of linear increases in RT to positive ($\mathrm{P\overline{D}}$) and negative ($\overline{\mathrm{PD}}$) probes as a function of $n(\mathrm{P} - \mathrm{D})$. Furthermore, when all D items are active (see left panel of Fig. 2), the slopes for reactions to $\mathrm{P\overline{D}}$ and $\overline{\mathrm{PD}}$ were approximately equal to each other and to the slopes observed for positive and negative reactions when no D items preceded the probe (see Fig. 1).

When all items in the D set are active, changes in $n(\mathrm{P} - \mathrm{D})$ and $n(\mathrm{D})$ are perfectly confounded and negatively correlated. Might not the above findings be interpreted as a simple consequence of changes in $n(\mathrm{D})$ instead of $n(\mathrm{P} - \mathrm{D})$? The answer is negative for two reasons. First, RT would not be expected to vary inversely with $n(\mathrm{D})$ (see left panel of Fig. 2). Second, when the two variables are no longer confounded because only inactive items are present in D, the two sets of functions relating RT to $n(\text{D})$ are strikingly different (see left and right panels of Fig. 2).

Additional support for a deletion strategy comes from the observation that when $n(\text{P} - \text{D})$ was constant at 4 (see right panel of Fig. 2), reactions to $\overline{\text{PD}}$ and $\overline{\text{P}}\text{D}$ were not significantly different. That is, negative reactions to D_I items were not reliably faster than reactions to nonpresented items. Had the two sets of curves differed significantly, the inference might have been drawn that the two types of items have unique representations in memory, contrary to a pure deletion notion.

One notable exception to a deletion interpretation is present in the data when homogeneous D sets are used. The curve labeled PD in the left panel of Fig. 2 initially increases as $n(\text{P} - \text{D})$ increases from 1 to 2, but it then drops when $n(\text{P} - \text{D}) = 3$. This single deviant data point is based on reactions to D_A items when D consisted of only one D_A item. Reactions to the single D_A item were quite fast. One possible explanation for this result is that the last D item is still "active" in memory and a probe can in some way be directly compared to it. This conjecture is not without limitation. It is apparent that when a single D_I element is the only member of D ($\overline{\text{P}}\text{D}$ curve in Fig. 2), there is little reduction in RT to a probe identical to a D_I element. Therefore, it may be best to limit the possibility of an "active" memory item to only D_A items.

With D sets containing only D_A items, the lack of a significant relationship between RT and the order correspondence in which identical elements were presented in P and D lends indirect support to a deletion interpretation. The values of the rank order correlations did not produce systematic differences in RT. One possible interpretation of this result may be associated with the 750-msec. delay between the last D item and the onset of the probe. The delay may have been greater than the time necessary to complete all deletions. If this is true, the effects of the order correlation may not have been manifest due to the length of the delay. Perhaps a shorter delay would allow an examination of this possibility and a better insight into the process of deletion. Even with the rather long interval and only D_A items in D, the average RT is about 100 msec. longer than those observed in the base-line condition with comparable $n(\text{P} - \text{D})$ sets. The result is comprehensible if one assumes that the longer the item resides in STM, the lower its probability of retrieval at the time of test. The error data agree with this possibility since error rates were much greater for both positive and negative probes when a D set was interpolated between P and the probe stimulus.

The data for D sets containing both D_A and D_I are much less consistent with a simple deletion model. One aspect of the data agreeing with a deletion notion comes from the results when $n(\text{P}) = 4$ and $n(\text{D}) = 3$. When RT was depicted as a function of $n(\text{P} - \text{D})$ (see top curve in Fig. 4), a significant linear component was observed. Here, $n(\text{D})$ remained constant, but the number of D_A items (items to be deleted) varied.

The presence in Fig. 4 of a significant quadratic component in the functions relating RT to $n(\text{P} - \text{D})$ precludes a simple interpretation of the data. If *S*s merely deleted items from P and compared the probe to the remaining items, functions relating RT to $n(\text{P} - \text{D})$ should not have been curvilinear. At a purely descriptive level, functions are curvilinear because *S*s experienced greater difficulty when D was composed of both active and inactive items. It is difficult to incorporate the elevation of RTs observed in these conditions into a deletion model unless a more complex strategy is considered. The *S*s reported that they knew after the D set was shown whether it contained all old (active) or all new (inactive) items; they also knew when D sets were mixtures of active and inactive items. This observation plus the nonlinear relationship between RT and $n(\text{P} - \text{D})$ suggests that *S*s may have employed a complex strategy involving conditional decisions based on the composition of the D set.

Alternative search models were applied to the data in an attempt to rationalize the curvilinearity. One such model assumed that the starting point for search depended on an initial detection of whether the D set was mixed (contained both D_A and D_I items) or unmixed. Suppose that deletions were performed and *S*s formed a $\text{P} - \text{D}$ set. At the same time, the D set was also available in memory. The *S*s may have had a decision rule of the form: if D is unmixed, compare the probe to $\text{P} - \text{D}$ items; and if D is mixed, compare the probe to the D elements. When D is searched first and if a match is made in D, a "no" response could be given. If a match was not made, *S* must examine $\text{P} - \text{D}$ to see if the test stimulus is contained within that set. Some aspects of the data are consistent with this strategy. The increases in RT to positive probes as a function of $n(\text{P} - \text{D})$ are predictable since the model assumes that P is reduced in size by deletion.

Also consonant with the model is the curvilinearity observed in the two upper curves in Fig. 4. The points associated with D sets containing active and inactive items are elevated (middle two points in top curve and middle point in the $n(\mathrm{P}) = 4$, $n(\mathrm{D}) = 2$ function). This elevation of positive reactions is understandable since a nonproductive search of D was conducted first. Another notable feature of the two middle data points in the uppermost curve in Fig. 4 is the fact that reactions were faster with two $\mathrm{D_A}$ items and one $\mathrm{D_I}$ item than with one $\mathrm{D_A}$ item and two $\mathrm{D_I}$ items. This result is expected since with mixed D sets, D is examined first. A positive probe could not be matched by a search through D; therefore, *S* is obligated to conduct a search of P − D. With $n(\mathrm{P} - \mathrm{D}) = 2$, a match would on the average be made sooner than with $n(\mathrm{P} - \mathrm{D}) = 3$. A final point concerns the overall elevation of the data for D-set conditions above the base-line condition. The model could account for this increase in vertical displacement by proposing that the initial decision of whether D is mixed or unmixed is a process that would add a constant amount of time to all D-set conditions.

This more complex model must be considered incomplete for one of the same reasons the simpler deletion model was insufficient. With all D items active, negative reactions to $\mathrm{D_A}$ probes should not have been faster than RTs to positive probes. However, in one instance, faster reactions were observed (with $n(\mathrm{P}) = 4$ and $\mathrm{D_A} = 1$ in left panel of Fig. 2). Reactions should have been faster to $\mathrm{P\overline{D}}$ items since these were examined first. A second reason for considering the present model unsatisfactory is that reactions to $\mathrm{D_I}$ items were sometimes slower than reactions to $\mathrm{P\overline{D}}$ items when D sets were mixed. These data were not described earlier, but in no instance should RTs to $\mathrm{D_I}$ items be faster than reactions to items in P − D when D was mixed.

In summary, when D sets contain either all active or all inactive elements, most features of the data agree with the proposition that *S*s delete items from P and search the P − D set. Difficulties of interpretation occur when D contains mixtures of both active and inactive items. A simple deletion model cannot account for all of the results. The possibility that *S*s used a more complex strategy based on the composition of D seems to provide a reasonable alternative, although it too has some notable limitations.

REFERENCES

Hilgard, E. R., & Bower, G. H. *Theories of learning.* New York: Appleton-Century-Crofts, 1966.

Posner, M. I. Short-term memory systems in human information processing. *Acta Psychologica,* 1967, **27**, 267–284.

Sternberg, S. High speed scanning in human memory. *Science,* 1966, **153**, 652–654.

Sternberg, S. Inferring the organization of cognitive events from reaction time: Some scanning processes in active memory. Paper presented at the symposium on research in cognitive psychology, City University of New York, June 1968.

(Received March 22, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 427–434

PATTERN RECOGNITION: DIFFERENCES BETWEEN MATCHING PATTERNS TO PATTERNS AND MATCHING DESCRIPTIONS TO PATTERNS[1]

GILLIAN COHEN[2]

Institute of Experimental Psychology, University of Oxford

Same–difference reaction times (RTs) for matching patterns to patterns (the PP condition) and for matching descriptions to the same patterns (the DP condition) were compared. Three levels of pattern complexity and three interstimulus intervals (ISIs) were employed. Where the RTs for the two conditions do not differ, it is argued that *S* may be converting the first stimulus from a description to a visual code to perform the matching task. RTs diverge with increased complexity and with lengthened interval. The DP condition is more sensitive to increased complexity, and the PP condition is more sensitive to lengthened interval.

Recognition involves matching a given stimulus to a representation of it formed and stored in memory when it was previously encountered. A number of facts have emerged from recent studies on letter recognition. Visual matching, i.e., a judgment of physical identity, is reliably faster than name matching (Posner & Mitchell, 1967). But the superiority of visual matching begins to fall off as the interval between the first and second presentations of the stimulus lengthens (Posner & Keele, 1967). It has also been established that matching is not necessarily carried out in the same mode as the first presentation of the stimulus. The form of the stored representation may be altered, and, consequently, the matching procedure employed may also be different. For instance, in some conditions, *S* can transform an auditorily presented letter name and generate from this a visual code producing recognition times equivalent to simple visual matching (Posner, Boies, Eichelman, & Taylor, 1969).

The relative efficiency of the two matching systems, visual and verbal, is affected by a number of factors. Time interval has been shown to be an important one. The generation of a visual code for a single auditorily presented letter takes approximately 750 msec.; on the other hand, the retention of unconverted visual coding seems more vulnerable to decay since the superiority of visual matching declines rapidly. With long interstimulus intervals (ISIs), therefore, *S* may be forced to store the first stimulus verbally as well and, if the visual code is too degenerate, to employ a name-matching strategy. It seems likely also that the complexity of the stimulus will affect both generation and retention.

The present experiment was an exploratory study designed to compare matching a pattern to a pattern with matching a visually presented written description to the pattern. In theory, *S* could convert the original stimulus from verbal information to a visual coding or vice versa. However, it is likely, as Posner et al. (1969) have suggested, that the ability to convert the original stimulus and retain it in the converted form may be facilitated or impeded by other factors. In the present study, the time interval between the original stimulus and the comparison stimulus was varied; and patterns, not letters, were used so that complexity could also be manipulated. The aim was to compare pattern–pattern matching (the PP condition) with description–pattern matching (the DP condition) and to discover how these are differentially affected at various levels of both factors.

---

[1] This work was carried out while the author was in receipt of a Medical Research Council scholarship.

[2] Requests for reprints should be sent to Gillian Cohen, Institute of Experimental Psychology, University of Oxford, 1 South Parks Road, Oxford, England.

Recent papers by Sekuler and Abrams (1968) and Nickerson (1967) give rise to the expectation that results for judgments of "same" will not follow the same trend as results for judgments of "different" since, they suggest, two different processes may be at work. Sekuler and Abrams found that *S*s visually matching matrices of filled cells produced judgments of "same" that were, up to a point, unaffected by increasing complexity or number of constituent elements in the matrix, although this was not the case for judgments of "different." From this data, they conclude that some form of gestalt, template, or parallel matching is possible in judgments of visual sameness. The present experiment also served to explore the limits of this ability.

METHOD

Materials.—The stimuli were made up as follows on 6 × 4 in. plain white cards. The patterns were of three types.

The low-complexity (LC) patterns varied on two dimensions, with three possible values in each. They consisted of a shape (triangle, circle, or square) colored red, blue, or yellow. The shapes were cut from gummed colored paper and mounted on the cards.

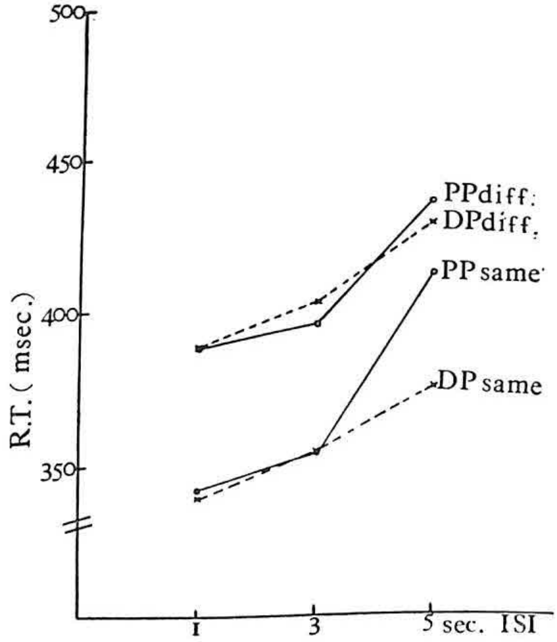


FIG. 1. Mean RTs for low-complexity patterns with ISIs of 1, 3, and 5 sec. (All figures show same and different judgments for both PP and DP matching.)

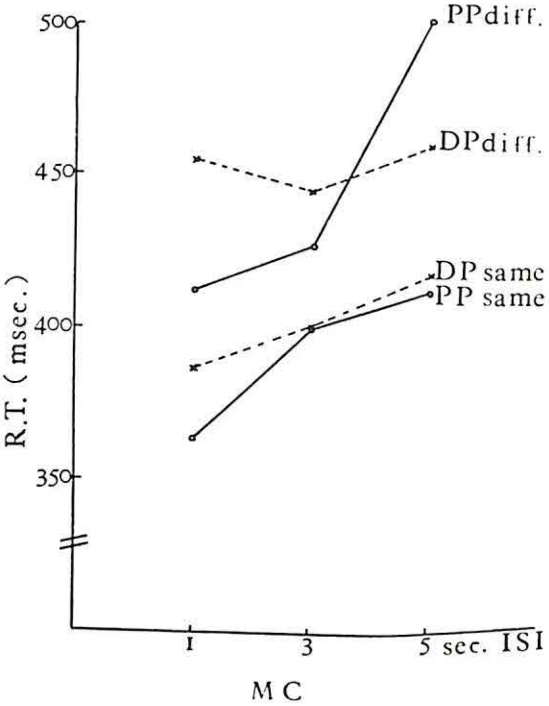


FIG. 2. Mean RTs for medium-complexity patterns with ISIs of 1, 3, and 5 sec.

The medium-complexity (MC) patterns had three potentially relevant dimensions or features: the shape, the color of the shape, as in LC, and in addition, the color of the spot which was always present in the center of the shape. The color of the spot was always different from the color of the main shape, but was again limited to the three values, red, blue, or yellow.

The high-complexity (HC) patterns had four relevant features: the shape, the color of the shape, and the color and orientation of a stripe that was placed across the shape. The stripe color again was either red, blue, or yellow, but different from the color of the main shape. The orientation could be either diagonal, horizontal, or vertical.

In the DP condition, the description was typed on the card. The following are examples: in the LC group, "a blue triangle"; in the MC group, "a red circle with a yellow spot"; and in the HC group, "a yellow square with a red vertical stripe."

When the second stimulus was "same," the pattern was identical with the first stimulus (PP) or exactly accorded with the description (DP). When the second stimulus was "different," only one feature was varied, but within each group each feature was varied equally often.

Subjects.—Eight *S*s completed the experiment. They were students from various disciplines, three

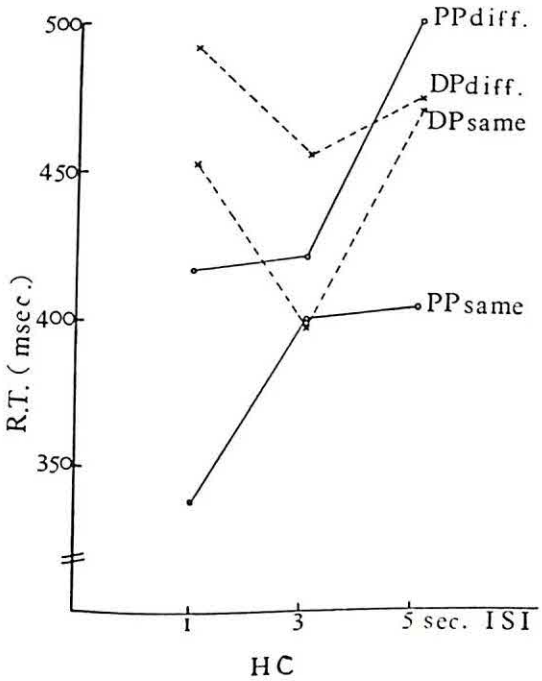


FIG. 3. Mean RTs for high-complexity patterns with ISIs of 1, 3, and 5 sec.

female and five male. All *S*s were unaware of the aims of the experiment and unfamiliar with the background. Each *S* was run individually in six sessions separated by a few days. They were paid 5 shillings per session.

Procedure.—The stimuli were viewed tachistoscopically. The patterns all subtended a visual angle of 2.7°. The typed descriptions subtended a visual angle of 6.5°, approximately. The first stimulus (whether pattern or description) was presented for 2 sec., which allowed ample time for reading even the HC descriptions. It was preceded by a verbal warning signal from *E*. The ISI was either 1, 3, or 5 sec. After the ISI, the second stimulus was automatically presented and remained present until *S* terminated it by responding. The onset of the second stimulus started a Venner timer, which was stopped by *S*'s response. The response required was a key press. Half of the *S*s pressed the "same" key with the dominant hand and the "different" key with the nondominant hand; half had the opposite arrangement. The *S* was instructed to judge the second stimulus same or different from the first and to press the appropriate key as fast as possible, but to try to avoid errors. At the beginning of a trial, *S* sat with his hands resting lightly on the response keys. Reaction times (RTs) and errors were recorded, and *S* was told his RT after each trial. The intertrial interval (ITI) was about 10 sec., and *S* rested briefly after each block of 18 trials. A block of 30 practice trials with a different set of practice cards preceded the experimental trials. Also at each subsequent session, *S* began with 12 practice trials, 4 each of the LC, MC, and HC types, with the ISI and matching condition (DP or PP) to be employed in that session. Each session lasted about 45 min.

Design.—Four *S*s had their first three sessions on the PP condition and their last three with the DP condition. The other four had the opposite order. At a single session, the ISI remained constant, but half of the *S*s worked from the shortest interval at Session 1 to the longest at Session 3 and, similarly, from shortest to longest in Sessions 4–6. Half of the *S*s worked in the other direction, starting each condition with the longest interval and proceeding to the shortest. Time intervals constant within sessions were used in preference to mixed time intervals since the effects of temporal uncertainty resulting from mixed time intervals would be confounded with the effects of the factors to be studied.

Within each session, *S* had three blocks of 18 trials, 54 in all. Half of the *S*s started with the LC type for the first block, proceeding through MC to HC, while half had the order HC, MC, LC. It was thus possible to produce an exactly balanced Latin-square design, in which the order of the two matching conditions (DP and PP) and the two dependent variables (time interval and complexity) was different for each of the eight *S*s. The aim was to eliminate the effects of practice as far as possible in the pooled data. The *S* was always informed at the beginning of a session of the timing, matching condition, and level of com-

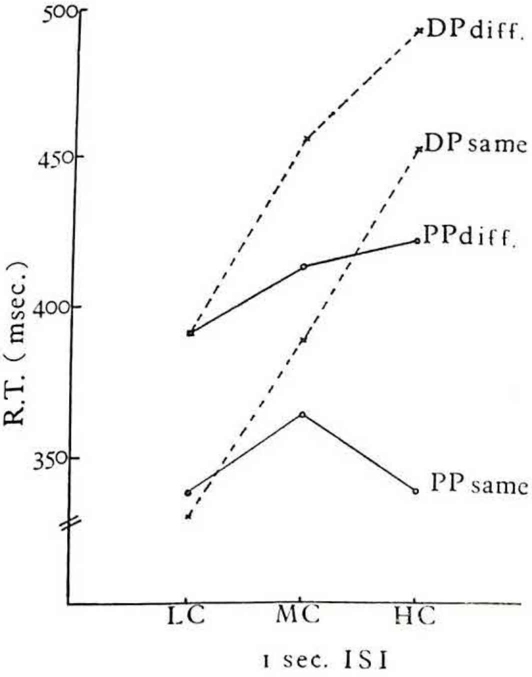


FIG. 4. Mean RTs for 1-sec. ISI with low-, medium-, and high-complexity patterns.

plexity that would confront him. He was thus given opportunity to adjust his strategy accordingly. The "same" and "different" trials were randomly mixed, with the constraint that no more than three of one kind were consecutive. The order of the features varied in the "different" trials was also randomized. The response entropies were equal.

RESULTS

The mean RTs for the eight *S*s are given in Table 1. Error scores were discarded from the data, and error trials were not replaced. Two four-way analyses of variance were performed, one for the "same" RTs and one for the "different" RTs. In each, the factors were *S*s, complexity, time interval, and matching condition. The results were significant.

In the "same" RTs, the interval factor was significant, $F\ (2, 28) = 14.04$, $p < .001$, and the complexity factor was also significant, $F\ (2, 28) = 23.43$, $p < .001$. The Matching Condition × Interval interaction was significant, $F\ (2, 28) = 6.3$, $p < .01$.

In the "different" RTs, the interval was significant, $F\ (2, 28) = 19.5$, $p < .001$. The Matching Condition × Complexity interaction was only just significant, $F\ (2, 28) = 3.53$, $p < .05$.

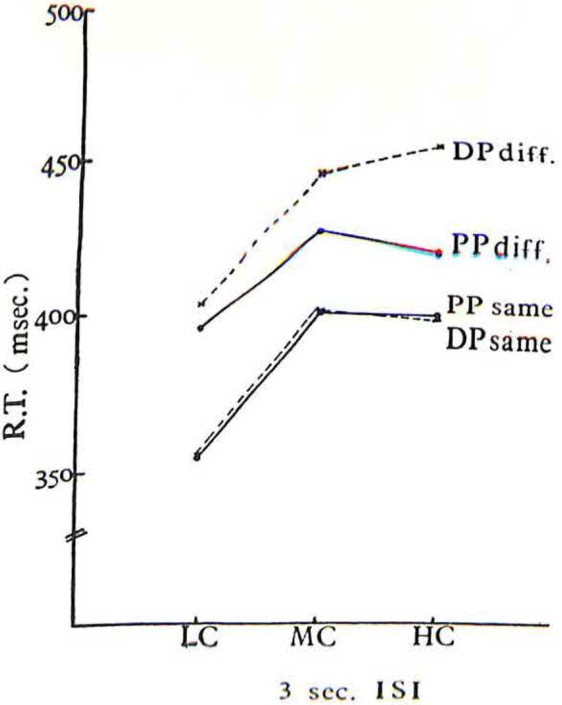


FIG. 5. Mean RTs for 3-sec. ISI, with low-, medium-, and high-complexity patterns.

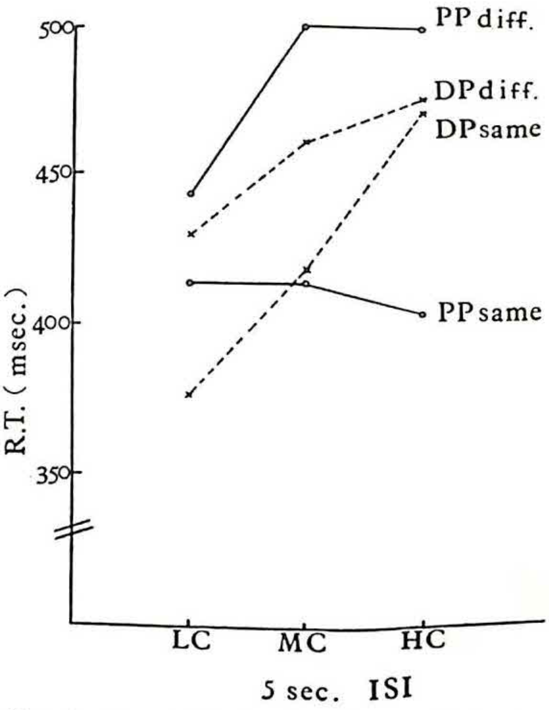


FIG. 6. Mean RTs for 5-sec. ISI, with low-, medium-, and high-complexity patterns.

The effect of the matching condition variable appears more clearly in a series of three-way analyses of variance. Separate analyses were carried out for the "same" RTs and for the "different" RTs at each of the three levels of complexity, HC, MC, and LC. In these, the factors were matching condition, interval, and *S*s. The PP and DP matching conditions differed significantly only in the "same" HC block, $F\ (2, 14) = 4.9$, $p < .05$. In the "same" LC analysis, there was a significant Matching Condition × Interval interaction, $F\ (2, 14) = 4.0$, $p < .05$. In the "different" RTs, the DP and PP conditions did not yield a significant difference due to the considerable variance. A further series of two-way analyses in which the factors were *S*s and matching condition showed that the "different" RTs did differ significantly in the HC 5-sec. group, $F\ (1, 14) = 26.4$, $p < .001$, and in the MC 5-sec. group, $F\ (1, 14) = 8.5$, $p < .025$. At other combinations of time interval and complexity, there was no difference. The Newman-Keuls test (Winer,

TABLE 1

Mean RTs with Standard Deviations Pooled across Eight *S*s

Cond.	Complexity								
	LC			MC			HC		
	1-sec. ISI	3-sec. ISI	5-sec. ISI	1-sec. ISI	3-sec. ISI	5-sec. ISI	1-sec. ISI	3-sec. ISI	5-sec. ISI
PP same									
M	342	353	413	363	401	413	342	401	404
SD	103	105	181	105	133	163	104	136	137
DP same									
M	336	354	376	386	402	420	452	398	472
SD	101	90	108	100	137	120	245	101	236
PP different									
M	389	395	443	412	427	505	417	421	503
SD	123	115	227	139	137	206	138	134	258
DP different									
M	389	404	430	455	447	461	491	455	476
SD	102	63	93	150	83	136	191	98	111

1962) was used to determine which pairs of means differed significantly. The results can be summarized as follows.

Differences between the Matching Conditions

"Same" RTs.—At HC the visual matching condition is superior at both short and long intervals, but not at the intermediate 3-sec. interval. At LC there is no difference at 1 sec. and at 3 sec., but the PP condition is worse at the 5-sec. interval. At MC the two matching conditions do not differ.

"Different" RTs.—In the "different" RTs, a crossover effect occurs. The PP condition is better at short intervals, but worse at long intervals at the MC and HC levels of complexity.

Differences within the Matching Conditions

The PP condition.—The "same" RTs increase with interval, but not with complexity. The "different" RTs increase with both interval and complexity.

The DP condition.—The "same" RTs increase with complexity, but the interval effect is slight. The "different" RTs increase with complexity, but not with interval.

These differences are all significant at the .01 level.

RTs with different features relevant

Mean RTs of "different" judgments in the HC blocks were compared for each of the features that could be relevant at each of the time intervals: shape, color of shape, color of stripe, and orientation of stripe. RTs vary according to the feature in which the pattern differs. In the DP condition, the order from fastest to slowest was color of shape, color of stripe, shape, and stripe orientation. This order remained stable across all three time intervals. In the PP condition, the order varied at each time interval.

Error Distribution

Both matching conditions showed some increase in errors at the HC level. The DP condition produced a higher proportion of false positives to false negatives (71:39) than the PP condition (51:45).

Practice Effects

These are shown in Table 4. Performance improved at each session except at Day 4 when all *S*s changed conditions.

The standard deviations from the means in all the data are large; this results from the occasionally very long RTs of over 1

TABLE 2

MEAN RTs WITH STANDARD DEVIATIONS FOR DIFFERENT FEATURES RELEVANT IN "DIFFERENT" JUDGMENTS OF HIGH-COMPLEXITY PATTERNS

Cond.	Relevant feature			
	Shape	Color	Stripe color	Stripe orientation
1-sec. ISI				
PP				
M	474	392	388	428
SD	173	138	97	160
DP				
M	504	387	454	557
SD	83	80	129	230
3-sec. ISI				
PP				
M	415	359	448	456
SD	135	60	104	202
DP				
M	468	401	457	480
SD	67	53	109	133
5-sec. ISI				
PP				
M	570	460	462	538
SD	347	183	248	232
DP				
M	492	449	452	548
SD	97	114	108	108

sec., which occur in the data of all *S*s. While these are most probably due to momentary lapses of attention and are discarded by some *E*s, they are retained here since there is no proof that their occurrence is irrelevant.

TABLE 4

MEAN RTs WITH STANDARD DEVIATIONS POOLED ACROSS *S*s FOR SUCCESSIVE DAYS

Statistic	Days					
	1	2	3	4	5	6
M	482	429	406	410	377	374
SD	204	129	126	166	97	218

DISCUSSION

Where the RTs for the two conditions do not diverge, two inferences are possible. Either the two systems, name matching and visual matching, yield similar performances or conversion has taken place in one condition, so that the same matching system is being employed with equal efficiency in both conditions. Conversely, where RTs for the two conditions diverge, two explanations are possible. Either the two matching systems are not equally fast or conversion is inadequate, so that, although both conditions employ the same matching procedure, efficiency is unequal. Clearly, the data of the present experiment do not provide conclusive evidence for either of these alternatives. However, some points of interest emerge from the results.

PP condition.—The "same" RTs increase with interval as the visual coding decays over time. The results favor parallel matching of features since the RTs do not increase with complexity. The fact that PP "same" RTs are less affected by the lengthened interval than PP "different" RTs, especially at HC (see Fig. 3), requires explanation. When the second stimulus is identical with the first, it may serve to reactivate the visual trace, even if this has begun to decay: when the second stimulus is dif-

TABLE 3

ERROR DISTRIBUTION

Cond.	Complexity								
	LC			MC			HC		
	1-sec. ISI	3-sec. ISI	5-sec. ISI	1-sec. ISI	3-sec. ISI	5-sec. ISI	1-sec. ISI	3-sec. ISI	5-sec. ISI
PP same	4	5	4	4	2	6	5	9	6
DP same	3	3	2	1	3	3	9	7	8
PP different	6	3	3	5	3	10	6	5	10
DP different	8	4	8	10	7	11	7	4	12

Note.—Each figure represents the number of errors out of a total of 72 trials.

ferent, this reactivation process is absent, and *S* must make his "different" judgment on the basis of the noisy or imprecise visual representation. The "different" RTs show that the more complex the visual coding, the less efficient it is for matching after some decay has taken place. The slope of the PP "different" scores plotted against interval becomes steeper in the MC and HC groups (see Fig. 2 and 3).

DP condition.—Three alternative explanations of the DP scores are possible:

1. The *S* always matches verbally. If this is so, it is hard to see why performance declines at the longest intervals since a verbal code may take time to organize, but once organized should remain stable. Also, on this interpretation it is necessary to assume that the internal verbalization must be highly abbreviated to produce RTs as fast as those obtained, and it must be reorganized in such a way as to reduce or obliterate the differences in number of syllables. Table 5 shows that RTs do not increase regularly with increases in the number of syllables in the description. The original description must also be reordered since Table 4 shows that the relative speed of RTs for different relevant features does not reflect the order of the features as presented in the description.

2. The *S* generates a visual code from the description and matches visually. On this interpretation, the V shape discernible in Fig. 2 and 3 could be taken as evidence that the generation of the more complex visual codes

TABLE 5

Mean RTs with Standard Deviations for DP Judgments with Descriptions Varying in Number of Syllables

Description	Complexity								
	LC			MC			HC		
	1-sec. ISI	3-sec. ISI	5-sec. ISI	1-sec. ISI	3-sec. ISI	5-sec. ISI	1-sec. ISI	3-sec. ISI	5-sec. ISI
2 syllables									
M	324	350	392						
SD	83	58	95						
3 syllables									
M	366	385	415						
SD	91	92	94						
4 syllables									
M	331	376	411						
SD	88	78	120						
5 syllables									
M	359	396	401	382	413	449			
SD	93	70	106	94	101	150			
6 syllables									
M				429	444	429			
SD				107	152	120			
7 syllables									
M				402	415	441			
SD				97	82	129			
8 syllables									
M							468	442	426
SD							235	196	103
9 syllables									
M							418	435	489
SD							111	90	140
10 syllables									
M							404	408	490
SD							122	94	240
11 syllables									
M							485	479	508
SD							105	169	128

reaches optimal efficiency at the intermediate time interval.

3. The *S* switches between strategies, matching verbally at some time–complexity combinations and visually at others.

Further work is obviously required to show which of these interpretations is correct. The results of the present experiment also differ in two respects from the findings of Posner and Keele (1967) and of Sekuler and Abrams (1968). The superiority of visual presentation does not break down until the ISI is between 3 and 5 sec., whereas Posner and Keele found that it reduced to zero by 1.5 sec. This difference may stem from the type of stimulus used here. The visual coding of a colored pattern may be more persistent than that of a letter, and the description of a pattern involving several words would be more difficult to convert than a letter name. The PP "same" RTs in the present experiment still remain constant when complexity is increased until the patterns contain four features. Sekuler and Abrams found that this constancy had already begun to break down when the number of filled cells in the matrix reached four. Again, the difference may be due to the type of stimulus. A colored pattern could be easier to treat as a whole than a number of cells scattered in a matrix. It may also be easier to employ a number of different feature analyzers in parallel than a single feature analyzer testing solely for location.

REFERENCES

NICKERSON, R. S. "Same-different" response times with multi-attribute stimulus differences. *Perceptual and Motor Skills,* 1967, **24**, 543–554.

POSNER, M., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. H. Retention of visual and name codes of single letters. *Journal of Experimental Psychology,* 1969, **79**(1, Pt. 2).

POSNER, M., & KEELE, S. W. Decay of visual information from a single letter. *Science,* 1967, **158**, 137–139.

POSNER, M., & MITCHELL, R. F. Chronometric analysis of classification. *Psychological Review,* 1967, **74**, 392–409.

SEKULER, R. W., & ABRAMS, M. Visual sameness: A choice time analysis of pattern recognition processes. *Journal of Experimental Psychology,* 1968, **77**, 232–238.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received March 26, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 435–440

STIMULUS PROBABILITY AS A DETERMINANT OF DISCRETE CHOICE REACTION TIME

HAROLD L. HAWKINS[1] AND KAREN HOSKING

University of South Florida

The present experiment was designed to test the hypothesis that stimulus probability is a critical determinant of choice reaction time (RT) only under conditions where the stimulus alternatives in question are represented within *S*s positive set. The positive set refers to the array of stored alternatives against which the stimulus input on each trial is compared prior to response selection. The likelihood that a particular pair of unequally probable stimuli would appear in *S*'s positive set was varied across conditions by manipulating the pattern of S-R correspondences. Mean RTs, error frequencies, and subjective reports obtained under these conditions tended to support the experimental hypothesis. The implications of the findings are discussed with regard to recent studies investigating the relative significance of stimulus and response probability in choice RT.

Beginning with the work of Donders (1868), many theoretical analyses of choice reaction time (RT) have held that two basic decision-making operations intervene between the stimulus and the response (e.g., Smith, 1968; Welford, 1960; Woodworth, 1938). The *S* must decide which of several possible stimuli has appeared (stimulus identification, or discrimination) and which of several possible responses is appropriate (response selection). However, the relative significance of these two operations in the determination of choice RT remains unclear. Consider, e.g., the well-known finding that choice RT varies as the inverse of the probability of alternatives when stimuli and responses are in 1:1 correspondence (Fitts, Peterson, & Wolpe, 1963; Leont'ev & Krinchik, 1963). Does this result imply that less time is necessary to identify frequently occurring stimuli or rather that less time is necessary to select frequently required responses? Two recent attempts to separate the effects of stimulus and response probability through the use of many:1 S-R correspondences offer contradictory answers. As part of a larger study, Dillon (1966) assigned a set of seven unequally probable alphabetic symbols to one of a pair of unequally probable key-press responses, and a set consisting of a single symbol to another. Choice RT was found to increase as the inverse of response probability, which varied across *S*s within the 1:1 S-R correspondence, but remained invariant across levels of stimulus probability within the many:1 mapping. Consequently, Dillon concluded that anticipatory adjustments in the response selection phase of information processing, not in the stimulus identification phase, critically determine RT. Bertelson and Tisseyre (1966) assigned two alphabetic symbols to each of two unequally probable key-press responses, with stimulus probability varying within the more probable of the two stimulus sets. Contrary to Dillon's findings, RT increased as the inverse of stimulus probability and remained constant across response probability values. On this basis, it was concluded that bias in the stimulus identification process critically determines choice RT.

Implicit in these and most other interpretations of choice RT data is the assumption that the anticipatory adjustments producing probability effects in RT are determined by the informational properties of the total S-R array provided to *S*. A study by Sternberg (1966) has shown that this assumption may be incorrect under certain conditions. Prior to each RT trial (Exp. I), or block of trials (Exp. II), *S* was presented with an array of items (positive set) drawn from the 10 digits. This set required one of two equiprobable lever-pull responses. All digits not prespecified (negative set) were assigned the other response. Under these conditions, RT increased with positive set size, but was uninfluenced by negative set size, suggesting that only the positive set had been maintained in memory and examined prior to response initiation. Responses to members of the negative set presumably were initiated when a match was not obtained between stimulus input and the stored set. Sternberg characterized the major independent variable in his study as set size. For present purposes, it is convenient to view this variable as stimulus probability. Thus, e.g., as the size of the positive set increased, the occurrence probability of any given alternative within this set declined. It should be emphasized that this alteration does not necessarily imply any further modification of Sternberg's interpretation.

---

[1] Requests for reprints should be sent to Harold L. Hawkins, Department of Psychology, University of South Florida, Tampa, Florida 33620.

This strategy, of selectively preattending only a portion of the total population of available relevant stimuli, is appropriate in many two-choice RT tasks, and in fact may be optimal because it minimizes the number of alternatives *S* must keep track of in memory and interrogate prior to response selection. In Dillon's (1966) task, e.g., it is plausible that *S*s stored the single symbol in 1:1 correspondence as the positive set, making the response appropriate for this stimulus on trials where it appeared and the alternative response on all other trials, which contained members of the (functionally negative) set assigned to the many:1 correspondence. Hence, Dillon's data may be interpreted as showing that choice RT is influenced by differences in the probabilities of the positive and negative sets, but not by differences in stimulus probability within the negative set. The first of these findings, viewed by Dillon as a response probability effect, is in fact equivocal in that stimulus and response probability were confounded within the positive set. The second finding is consistent with Sternberg (1966).

The occurrence of stimulus probability effects in Bertelson and Tisseyre's (1966) study may be related to the presence in both stimulus sets of multiple alternatives. Under these conditions there is no obvious reason why *S*s would consistently choose one or the other of the two sets to selectively rehearse. Thus, it seems likely that the set within which stimulus probability varied would sometimes serve as positive. Viewed in this way, Bertelson and Tisseyre's results, like those of Sternberg (1966), demonstrate that variations in stimulus probability within the positive set are an important determinant of choice RT.

In sum, the previous considerations imply that variations in stimulus probability within a particular stimulus set, A, will influence RT provided A is treated as positive; i.e., provided the alternatives constituting A are included among the array of stored items with which the stimulus presented on each trial is compared. However, if A is treated as negative and reactions to members of this set occur when a match is not obtained between stimulus input and the stored positive set, then RT will not be influenced by variations in stimulus probability within A. In other words, if the identities of the individual alternatives constituting A are not preserved, the probabilities of these alternatives should not be preserved nor subsequently reflected in RT performance.

Several implications of the foregoing analysis were examined in the present experiment. Three conditions were studied which should differentially influence the likelihood that a set, A, of unequally probable stimuli would be treated by *S*s as positive. Under Cond. 4-1, four unequally probable stimuli (Set A) were assigned one key-press response, and a single stimulus (Set B) was assigned another. Under this condition, which is similar to that studied by Dillon (1966), it was expected that B, not A, would serve as the positive set. Under Cond. 2-1-2, two unequally probable stimuli (A) were assigned one response, one stimulus (B) was assigned another, and two stimuli (NR) were assigned to a no-response category; i.e., the appearance of either stimulus signaled *S* to withhold a response. Here, B cannot serve as the positive set by itself as was possible under Cond. 4-1. In fact, accurate performance in Cond. 2-1-2 logically requires that at least two of the three stimulus sets are kept track of in memory. For this reason, and because A was one of the sets demanding an active response, it was expected that at least some *S*s would treat this set as positive. Under Cond. 2-3, two unequally probable stimuli (A) were assigned one response, and three stimuli were assigned the other. This condition is similar to that of Bertelson and Tisseyre (1966) in that both stimulus sets contained multiple alternatives. Accordingly, it was expected that here too at least some *S*s would treat A as positive. These suppositions led to the prediction that stimulus probability effects would appear within A under Cond. 2-1-2 and 2-3, but not under Cond. 4-1.

METHOD

Subjects.—The *S*s were 18 advanced undergraduate psychology students at the University of South Florida, who received course credit for participation in the experiment.

Stimulus materials and apparatus.—The stimulus population consisted of the upper case letters, D, H, Q, S, and X (Letraset No. 193). The letters were presented individually on a uniform white background at a viewing distance of 17 in. through a Polymetric two-field mirror tachistoscope. Reaction time was measured by a Hunter Klockounter.

Procedure.—Six *S*s were assigned to each of the three experimental conditions. Under Cond. 4-1, the four stimuli comprising Set A and requiring one of the two key-press responses were presented with the respective probabilities .40, .15, .15, and .05. The single stimulus comprising Set B and requiring the alternative response occurred with probability .25. Under Cond. 2-1-2, the two stimuli constituting Set A occurred with probabilities .40 and .05, the single stimulus defining B with probability .25, and the two no-response (NR) stimuli occurred equally often (.15). Under Cond. 2-3, the Set A probabilities were .40 and .05, and the Set B probabilities were .25, .15, and .15. The identity, number, and occurrence probabilities of stimuli were constant across conditions. In addition the two occurrence probabilities, .40 and .05, appeared in Set A under all conditions. Within conditions, the particular probability and stimulus set assignments of the five stimulus letters were determined randomly. Set A was assigned equally often with the dominant and nondominant hands across *S*s within each condition.

At the beginning of the first test session, *S*s were given 50 practice trials on a two-choice task with four equiprobable digits as stimuli. The digits 3 and 7 were assigned one response and 5 and 8, the other. The *S*s were then provided with experimental stimulus and response probabilities, and there followed 200 trials on each of 2 consecutive days of practice. A 2-min. rest interval was given following each block of 25 trials. At the conclusion of a 6-sec. intertrial interval, each trial was initiated by a warning tone of 1-sec. duration whose offset coincided with stimulus onset. After each correct response, *S* was informed of his RT in milliseconds. Following each incorrect response, *S* was told only that he had made an error. Stimuli to which *S* incorrectly responded were subsequently repeated within the same trial block.

RESULTS

Table 1 provides mean RT ($\overline{\mathrm{RT}}$) and percentage of incorrect responses for practice trials and for each stimulus probability value under the three experimental conditions. A comparison of $\overline{\mathrm{RT}}$s for the two common probability values, .40 and .05, was of principal theoretical interest. Consequently, a two-way analysis of variance was performed with stimulus probability (.40 and .05) as a within-*S*s factor and conditions as a between-*S*s factor. Condition, $F\ (2, 15) = 8.12$, $p < .005$, stimulus probability, $F\ (1, 15) = 102.8$, $p < .001$, and the interaction of these, $F\ (2, 15) = 23.18$, $p < .001$, were all significant components of variance. To assess the prediction that stimulus probability effects would appear under Cond. 2-1-2 and 2-3, but not under Cond. 4-1, a simple main effects analysis was carried out. This analysis indicated that choice RT increased as the inverse of stimulus probability under Cond. 2-1-2, $F\ (1, 15) = 73.3$, $p < .001$, and 2-3, $F\ (1, 15) = 75.7$, $p < .001$, but remained invariant under Cond.

TABLE 1

MEAN RT AND PERCENTAGE OF INCORRECT RESPONSES (IN PARENTHESES) FOR THE COMMON PRACTICE TASK AND FOR EACH STIMULUS PROBABILITY VALUE UNDER THE THREE EXPERIMENTAL CONDITIONS

Cond.	Practice task	Experimental task				
		Set A			Set B	
		.40	.15	.05	.25	.15
4-1	470 (8.7)	394 (1.4)	382 (1.0)	390 (1.7)	418 (15.2)	—
2-1-2[a]	487 (8.0)	440 (1.9)	—	575 (6.7)	445 (3.0)	—
2-3	406 (6.3)	342 (0.3)	—	487 (23.5)	384 (3.3)	423 (4.0)

[a] Percentage of errors to Set NR under Cond. 2-1-2 was 2.5.

4-1 ($F < 1.0$). If no response is considered a type of response, it could be argued that the number of response alternatives increased between Cond. 2-3 and 2-1-2. Because this variable was of potential theoretical interest, two further analyses were carried out on the data from these two conditions. The first was an analysis of variance identical with the previous except that the condition factor contained only the levels 2-1-2 and 2-3. Here, the Condition × Stimulus Probability interaction was nonsignificant ($F < 1.0$). The second, an analysis of covariance, was conducted to determine whether the large overall difference in $\overline{\mathrm{RT}}$ observed between conditions was due to task or to *S* characteristics. The covariate was $\overline{\mathrm{RT}}$ on the common practice task given prior to experimental trials. The condition effect disappeared in this analysis ($F < 1.0$), indicating that the observed differences were due to *S*, rather than task, characteristics.

A further mixed-model analysis of variance was performed on an arc sine transformation of the error data for the stimulus probabilities .40 and .05. Condition, F (2, 15) = 14.9, stimulus probability, F (1, 15) = 35.8, and the interaction of these, F (2, 15) = 19.8, were all significant sources of variance ($p < .001$). A simple main effects analysis indicated that error rate, like RT, increased as the inverse of stimulus probability under Cond. 2-1-2, F (1, 15) = 4.68, $p < .05$, and Cond. 2-3, F (1, 15) = 72.3, $p < .001$, but remained relatively invariant under Cond. 4-1 ($F < 1.0$). In addition, within Cond. 4-1 a reliable difference in error rate was obtained between stimulus Sets A (pooled average) and B, t (5) = 4.62, $p < .01$.

DISCUSSION

The present experiment has clearly demonstrated conditions under which stimulus probability both will and will not affect choice RT. However, these conditions require further explication.

Earlier in the present paper it was proposed that variations in stimulus probability within a particular stimulus set, A, will affect RT provided this set is included among the items represented in memory and examined prior to the selection of a response (positive set). It was further proposed that the likelihood of A serving as the positive set is less in two-choice tasks where the alternative stimulus set, B, contains a single alternative (e.g., Dillon, 1966; and Cond. 4–1) than in tasks containing more than two stimulus sets (e.g., Cond. 2-1-2), or two-choice tasks where both stimulus sets contain multiple alternatives (e.g., Bertelson & Tisseyre, 1966; and Cond. 2-3). Consistent support for these propositions appeared not only in the RT data, but also, and more directly, in the subjective reports elicited following the experiment and in the error data. First of all, it is noteworthy that all *S*s in all conditions indicated they had rehearsed (i.e., preattended) only a portion of the total population of available stimulus alternatives. To be more specific, *S*s tested under Cond. 4-1 stated without exception that they had rehearsed only the single stimulus defining Set B and had responded on each trial strictly on the basis of whether or not this alternative had appeared. All *S*s tested under Cond. 2-1-2 indicated they had rehearsed both Sets A and B, while ignoring the NR set. The *S*s in Cond. 2-3 can be divided into two subgroups. Three of the *S*s indicated that they had rehearsed only Set A throughout the experiment, and the remaining three indicated they had varied their strategies, on some trials rehearsing A only, and on others, B only. In line with these reports, the difference in $\overline{\mathrm{RT}}$ to the stimulus alternatives in Set B with probabilities .25 and .15 was 19 msec. for the first subgroup and 54 msec. for the second. This difference was statistically reliable, t (4) = 2.92, $p < .05$.

The error data suggested much the same picture. Under Cond. 4-1, error frequency was unrelated to stimulus probability within A. On the contrary, the distinguishing feature of the error pattern here was the difference between Sets A and B. In accord with the interpretation of Dillon's (1966) results proposed earlier, this difference indicates the presence during information processing of anticipatory adjustments based on the relative probabilities of the positive and negative sets. However, under Cond. 2-1-2 and 2-3, error frequency, like $\overline{\mathrm{RT}}$, is inversely related to stimulus probability within A, suggesting that anticipatory adjustments were based on the informational properties of stimuli in Set A.

Several objections might be raised concerning the previous formulation. First, if *S*s in Dillon's (1966) study and in Cond. 4-1 were in fact selectively rehearsing the single-stimulus alternative occurring in 1:1 correspondence, why is RT for this item generally longer than for all remaining alternatives? The logic of this argument is that selective rehearsal implies selective anticipation, and reactions to anticipated alternatives should generally be shortest. The problem here lies in the interpretation of selective rehearsal. Although *S* may rehearse a particular item prior to a given trial, he may not necessarily expect, or anticipate, that this item is likely to occur. In fact, since the probability of the alternative in 1:1 correspondence was less than for the remaining alternatives in both the Dillon study and Cond. 4-1, *S*s probably anticipated that the single alternative would *not* occur. A second objection might be that the interaction observed between Cond. 4-1 and 2-1-2 could have resulted from an increase in number of response alternatives rather than from changes in rehearsal strategies. While this interpretation is plausible, it does not account for the equivalent interaction obtained between Cond. 4-1 and 2-3. When *S* characteristics were taken into consideration, the obtained similarity of results between Cond. 2-1-2 and 2-3 indicated rather that the number of response alternatives had no measurable effect in the present experiment. A final objection follows a study by Laberge and Tweedy (1964) in which two unequally probable stimuli were assigned one of two unequally probable responses and a single stimulus was assigned the other. Although this task is similar to that used by Dillon and in Cond. 4-1, contrary to these, Laberge and Tweedy reported both stimulus and response probability effects. In terms of the present analysis, these results imply that *S*s did not consistently treat the single-stimulus alternative as the positive set. The same type of apparent nonoptimality occurred in Cond. 2-3, where several *S*s at times employed the larger set as positive. With specific regard to two-choice tasks, these findings suggest that the likelihood that *S*s will consistently treat the smaller of two stimulus sets as positive increases with the size imbalance obtaining between the sets.

In conclusion, the present experiment has shown conditions under which stimulus probability, and thence the stimulus identification process, is an important determinant of choice RT. Considering these results together with those reported by Bertelson and Tisseyre (1966), a stronger conclusion might be that response probability is unimportant in speeded choice tasks. However, data are available which suggest that such a generalization is unwarranted. A case in point is the literature on serial choice RT where the interval between the response on Trial j and stimulus onset on Trial $j+1$ is relatively brief (less than 2 sec.). Here, Kornblum (1968) has demonstrated a relationship between the probability of an S-R alternative and RT that is primarily the result of repetition effects. Bertelson (1965) has shown further that response, rather than stimulus, repetition is the principal factor in such tasks. Thus, response probability effects produced by response repetition are to be expected in serial choice RT (e.g., Laberge, Legrand, & Hobbie, 1969). It is noteworthy in this connection that neither stimulus nor response repetition effects were observed in the present data. That is, no systematic differences in RT were obtained between repeated and nonrepeated stimuli or responses under any of the three conditions studied. Finally, significant main and interaction effects for stimulus and response probability have been found by Pollack (1963) in a word-categorization task and by Rabbit (1959) in a card-sorting task. These differ in several major respects from the discrete choice RT experiments previously discussed in the present paper. First, both studies investigated serial rather than discrete reaction tasks. Second, response probability was manipulated by varying the number of equally probable responses rather than by varying relative frequencies. Lastly, the number of response alternatives ranged well above two. The possible implications of the first difference have already been mentioned. Rabbitt (1967) has raised an interesting point that may apply, by extension, to the second and third differences. Consider a choice task containing eight stimulus alternatives, A, B, C, D, E, F, G, and H, where two stimuli are paired with each of four responses, e.g., *AB/CD/EF/GH*. Here, a theoretically optimal discrimination strategy would use only those stimulus attributes necessary to distinguish stimuli within each set from stimuli in other sets. In this way, A need be distinguished from C, D, E, F, G, and H, but not from B, which requires the same response as A. Now suppose that four of the letters are paired with each of two responses (e.g., *ABCD/EFGH*). Here A need be distinguished only from alternatives E, F, G, and H, but not from B, C, or D. Hence, increasing the number of response alternatives, while holding the total number of stimuli constant, in theory may lead to an increase in the number of stimuli that must be discriminated. Increases in choice time between these two situations may, therefore, reflect an increase in the difficulty of stimulus discrimination rather than response selection.

REFERENCES

Bertelson, P. Serial choice reaction-time as a function of response versus signal-and-response repetition. *Nature,* 1965, **208**, 217–218.

Bertelson, P., & Tisseyre, F. Choice reaction time as a function of stimulus versus response relative frequency of occurrence. *Nature,* 1966, **212**, 1069–1070.

Dillon, P. J. Stimulus versus response decisions as determinants of the relative frequency effect in disjunctive reaction time performance. *Journal of Experimental Psychology,* 1966, **71**, 321–330.

Donders, F. E. Die Schnelligkeit Psychischer Processe. *Archiv Anatomie und Physiologie,* 1868, 657–681.

Fitts, P. M., Peterson, J. R., & Wolpe, G. Cognitive aspects of information processing: II. Adjustments to stimulus redundancy. *Journal of Experimental Psychology,* 1963, **65**, 423–432.

Kornblum, S. Serial-choice reaction time: Inadequacies of the information hypothesis. *Science,* 1968, **159**, 432–434.

LaBerge, D., & Tweedy, J. R. Presentation probability and choice time. *Journal of Experimental Psychology,* 1964, **68**, 477–481.

LaBerge, D., Legrand, R., & Hobbie, R. K. Functional identification of perceptual and response biases in choice reaction time. *Journal of Experimental Psychology,* 1969, **79**, 295–299.

Leont'ev, A. N., & Krinchik, C. P. On certain aspects of human information processing. *Soviet Psychology and Psychiatry,* 1963, **1**, 24–31.

Pollack, I. Speed of classification of words into superordinate categories. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 159–165.

Rabbitt, P. M. A. Effects of independent variations in stimulus and response probability. *Nature,* 1959, **183**, 1212.

Rabbitt, P. M. A. Learning to ignore irrelevant information. *American Journal of Psychology,* 1967, **80**, 1–13.

Smith, E. E. Choice reaction time: An analysis of the major theoretical position. *Psychological Bulletin,* 1968, **69**, 77–110.

Sternberg, S. High-speed scanning in human memory. *Science,* 1966, **153**, 652–654.

Welford, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics,* 1960, **3**, 189–229.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 27, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 441–447

INVARIANCE OF TOTAL LEARNING TIME UNDER DIFFERENT CONDITIONS OF PRACTICE[1]

ROSE T. ZACKS[2]

Temple University

The phrase "total-time invariance" (TTI) refers to the finding that the amount learned in a given interval of time is invariant, regardless of how that time is distributed. The present study demonstrated that the phenomenon of TTI has a high degree of generality. The amount of control which *S*s had over the temporal distribution of their activities was varied by allowing them to (*a*) study each pair as long as they wished, (*b*) take test trials whenever they chose to, (*c*) both, or (*d*) neither. These variations did not affect total learning time to criterion. In addition, the data provided evidence for a distinction between the "nominal" and the "covert" distribution of study time. In other words, even though *E* presents all the items for the same amount of time, *S* may in fact distribute his study time unequally over the items in the list. An hypothesis to account for TTI was advanced on the basis of this distinction.

Evidence from a variety of sources strongly indicates that the amount learned in a given interval of study time is invariant, regardless of how that time is distributed. Variation of presentation rate and manipulation of what may be broadly termed "method of practice" (e.g., part vs. whole learning, self- vs. *E*-pacing) usually do not affect performance at the end of a fixed amount of learning time. This experimental result is referred to as total-time invariance (TTI). An example is the finding that the total time (Time per Trial × Number of Trials) required to learn a list of items to some specified criterion is unaffected by large variations in presentation rate (Bugelski, 1962; Fischer, 1966). Cooper and Pantle (1967) have recently reviewed the relevant data.

The present study had two major purposes. The first was to investigate the generality of TTI. With few exceptions (e.g., Bugelski & Rickwood, 1963; Morsund & Chape, 1966), the experiments demonstrating TTI have employed learning procedures in which the temporal sequence of events is almost completely determined by *E*. The question arises as to whether total learning time will remain invariant as *S*s are given increasing control over the temporal distribution of their activities. In the present experiment, the degree of constraint imposed by the experimental arrangements on the distribution of study and test time was systematically varied in an attempt to answer this question. In the control condition (standard study–test method of paired-associate learning), study and test trials alternated and the items were presented at a fixed rate on both study and test trials. The three experimental treatments differed from the control condition in that *S*s could either (*a*) study each pair for as long as they wished on study trials, (*b*) take test trials whenever they chose to, or (*c*) both.

The second major purpose of the present experiment was to obtain certain information about how *S*s distribute their study time over the items in a list. It was expected that *S*s in the self-paced study conditions would study hard pairs for longer durations than easy pairs. However, even

---

[1] This article was based on a dissertation submitted to the graduate school of the University of California, Berkeley, in partial fulfillment of the requirements for the PhD degree. The author gratefully acknowledges the invaluable assistance of Leo Postman, who served as chairman of the dissertation committee. Geoffrey Keppel and William Rohwer, who also served on the committee, deserve thanks for their many worthwhile suggestions. The research was supported by a grant-in-aid from the Committee on Research, University of California, Berkeley, and by Research Grant MH-12006 from the National Institute of Mental Health to Postman.

[2] Requests for reprints should be sent to Rose T. Zacks, Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.

in standard verbal learning procedures in which all items are presented for equal durations on each trial, *S*s may not rehearse all items for equal amounts of time. They may concentrate on a few of the items during a particular presentation, as Waugh (1961) has suggested, or they may spend more time on items they find difficult than on those they find easy. It is argued that variation of presentation rate or of other aspects of procedure often does not affect the amount learned in a specified interval of time, as least partly because the manner in which *S*s covertly distribute their study time is little affected by procedural manipulations such as variation of presentation rate. Information about how *S*s actually distribute their study time was obtained by analyzing the study durations used by *S*s in the self-paced conditions and by questioning *S*s in the *E*-paced conditions.

Each *S* learned two lists of pairs so that transfer might be studied. The two lists were learned either under the same or different methods of practice. It was thus possible to study differential transfer from one method of practice to another.

METHOD

Design.—All *S*s learned two lists of paired associates which conformed to an A-B, C-D transfer paradigm. There were four conditions of first-list learning comprising a 2 × 2 factorial design. The two independent variables were time per pair on study trials—constrained (E-paced) vs. free (self-paced)—and order of study and test trials—constrained (*E*-controlled) vs. free (*S*-controlled). The four conditions of first-list learning are referred to as: CC—control condition, with time per pair on study trials and order of study and test trials both constrained; FC—time per pair on study trials free, order of study and test trials constrained; CF—time per pair on study trials constrained, order of study and test trials free; and FF—time per pair on study trials and order of study and test trials both free. Attainment of the criterion of first-list learning was followed by learning of the transfer list under the CC or FF procedure. With four conditions of first-list practice and two conditions of second-list practice, there were eight independent groups of *S*s, which are designated CC-CC, FC-CC, CF-CC, FF-CC, CC-FF, FC-FF, CF-FF, and FF-FF. (In each case, the first two letters refer to the condition of first-list acquisition and the last two letters refer to the condition of transfer-list acquisition.)

Materials.—The learning materials were lists of eight paired associates, with high-frequency one-syllable nouns as stimuli and low-frequency bigrams as responses. Two different lists of paired associates were used. The bigrams were selected on the basis of the bigram frequency count of Underwood and Schulz (1960, Appendix D). The mean frequencies of the response terms in the two lists were 3.4 (range of 0–18) and 3.3 (range of 0–14), respectively. Intralist and interlist formal similarity were minimized as much as possible for the response terms. Within each list, there were no repeated letters among the response terms; and across the two lists, nine letters were used twice, but in different positions. Similarity was also minimized for the stimulus terms. All 16 stimuli had different initial letters, and there were no obvious conceptual or associative relations among the stimuli within each list or across lists. Stimuli and responses were paired randomly except that the initial and final letters of the stimulus could not be the same as either letter of the response.

Each list was used equally often as the first or the transfer list. For each list there were eight presentation orders for both the study and the test trials. Starting orders for the study and test trials of each list were assigned at random.

Apparatus.—Two Kodak Carousel slide projectors were used to display the items on a projection screen in a semi-darkened room, one for study trials and the other for test trials. In addition to a timer which allowed for automatic operation of the projectors at a 2-sec. rate, there was a manual button which *S*s in the self-paced study conditions used to change slides at their own rate. The intervals at which *S*s changed slides were recorded automatically by an Esterline-Angus pen recorder. The paper was run at a speed of 2 ft/min, making it possible to measure the time spent studying each pair to the nearest .1 sec.

Procedure.—All *S*s were run individually. Prior to the beginning of first-list learning, all *S*s were read general instructions for paired-associate learning. They then received one study and one test trial at a 2-sec. rate on the appropriate first list. All *S*s were given an initial study and test trial under the same procedure in order to have some measure of the equality of learning ability among the groups. Following the initial study and test trial, *S*s were given further instructions appropriate to their condition of first-list learning. For the CC condition, a 2-sec. rate was used for study and test trials, and study and test trials alternated. For the FC condition, *S*s controlled the study time per pair on the study trials by pushing the manual button, but study and test trials alternated. In the CF procedure, a 2-sec. rate was used for both study and test trials, but *S* controlled the order of study and test trials. During each intertrial interval (ITI), *S* had to say whether he wanted the next trial to be a study trial or a test trial. In the FF procedure, *S* controlled both the study time per pair on study trials and the sequence of study and test trials. In all conditions, a 2-sec. rate was used for test trials, and the ITIs (which were timed by a stopwatch) were 10 sec. The interval between the initial study and test trial and the continuation of first-list learning was 2 min. The *S*s were required to spell the bigram responses out loud on test trials. They were instructed that the criterion of learning was one perfect recitation of the list. The *S*s were urged to try to learn the list as quickly as possible. In an attempt to prevent overlearning in Cond. FC, CF, and FF, *S*s in these conditions were told not to try to learn the whole list at once and not to avoid test trials, as test trials would probably help them to learn the list.

The procedure for the learning of the second list was the same as that for first-list acquisition. An initial study and test trial at a 2-sec. rate was followed by further instructions appropriate to the transfer condition. There were, however, only two transfer conditions, Cond. CC and FF. The interval between the end of first-list learning and the initial study trial of List 2 was 2 min. Ninety seconds elapsed between the end of the initial study and test trial and the continuation of transfer-list practice. List 2 was also learned to a criterion of one errorless recitation.

After completing the second list, *S*s in Cond. CC and CF were asked if they had rehearsed all pairs for equal amounts of time and, if not, which pairs they had rehearsed for more time and when.

Subjects.—Twenty University of California undergraduates served as *S*s in each of eight groups. All were naive to rote learning experiments. The experiment was run in blocks of eight *S*s, with one *S* from each group in each block. Running orders within blocks were determined by a table of random numbers. Three *S*s, each from a different group, were eliminated for failure to learn the first list within 40 min.

RESULTS

Performance on the initial test trial of List 1.—In all conditions, practice on the first list began with an initial study and test trial under the standard procedure. Performance on this test trial was scored by two methods. In the stringent scoring method, *S* was given credit for a correct response term only if it was paired with the appropriate stimulus. In the lenient scoring method, *S* was given credit for all different correct responses regardless of whether or not they were paired with the appropriate stimuli. Analysis of both the stringent and the lenient scores revealed only one significant difference among the groups. The groups assigned to the second-list FF condition tended to be superior to the groups assigned to the second-list CC condition. For the lenient scores, this difference was reliable. The mean lenient scores for the groups learning List 2 under the CC and FF procedures were 1.55 and 2.01, respectively, $F\ (1,\ 152) = 4.82$, $p < .05$. Also, within-group correlations between performance on the initial test trial of List 1 (lenient scores) and total learning time (excluding ITIs) on List 1 indicate that performance on the first trial is predictive of the dependent measures of the present experiment. The eight correlations ranged from $-.219$ to $-.490$, with a mean of $-.334$. Only two correlations attained significance at the .05 level, but all were in the expected direction; i.e., total learning time decreased as number correct on the initial test trial increased. Thus, it appears that the groups assigned to the List 2, FF condition were of superior ability to those assigned to the List 2, CC condition. It should be noted, however, that when the two groups learning the first list under the same procedure are combined (i.e., Groups CC-CC and CC-FF, Groups FC-CC and FC-FF, etc.), there is no evidence in the performance on the first trial of ability differences across the four conditions of first-list learning. For both stringent and lenient scores, the *F* ratios for List 1 condition and for the List 1 × List 2 interaction were less than 1.00.

Total learning times on List 1.—There are two total-learning-time measures, study-plus-test time and total time including ITIs. Since analysis of study-plus-test times and of total times including ITIs yielded equivalent conclusions in all instances, only the analyses of the study-plus-test times are reported. Also, study-plus-test time is often referred to simply as total time or total learning time; time scores include ITIs only when these are explicitly mentioned.

Table 1 presents the means of the time scores for all eight groups separately and for the four combined groups. Bartlett's test revealed that there was heterogeneity of variance among the four combined groups in the study-plus-test scores, $\chi^2\ (3) = 15.50$, $p < .01$. A log transformation eliminated the heterogeneity of variance; therefore, all analyses were performed on log transforms of the data.

Inspection of Table 1 indicates that study-plus-test times are quite similar for Cond. CC, FC, and FF and that learning under

TABLE 1

Means of Various Measures of List 1 Learning

Group	Study time	Test time	Study +test time	Total time—including ITIs
CC-CC	156.8	156.8	313.6	489.6
CC-FF	144.8	144.8	289.6	450.6
M Cond. CC	150.8	150.8	301.6	470.1
FC-CC	273.8	144.8	418.6	579.6
FC-FF	197.9	94.4	292.3	390.3
M Cond. FC	235.8	119.6	355.4	485.0
CF-CC	239.2	156.0	395.2	622.2
CF-FF	230.4	206.4	436.8	689.8
M Cond. CF	234.8	181.2	416.0	656.0
FF-CC	339.6	143.2	482.8	648.2
FF-FF	151.2	80.0	231.2	313.8
M Cond. FF	245.4	111.6	357.0	481.0

Note.—Times are given in seconds.

TABLE 2

Means of Various Measures of List 2 Learning

Group	Study time	Test time	Study +test time	Total time—including ITIs
CC-CC	112.8	112.8	225.6	346.6
FC-CC	103.2	103.2	206.4	315.4
CF-CC	120.0	120.0	240.0	370.0
FF-CC	108.0	108.0	216.0	331.0
M Cond. CC	111.0	111.0	222.0	340.8
CC-FF	124.2	81.6	205.8	393.4
FC-FF	120.5	59.2	179.7	238.7
CF-FF	187.7	97.6	285.3	399.8
FF-FF	106.6	55.2	161.8	214.8
M Cond. FF	134.8	73.4	208.2	286.7

Note.—Times are given in seconds.

Cond. CF is slightly slower than under the other three conditions. However, analysis of variance shows that the differences in total learning times are not significant, all Fs $(1, 156) \lesseqgtr 2.32$, $p > .05$.

Constancy of total learning times across conditions does not preclude the possibility of differences in the form of the learning curves. However, an inspection of the List 1 learning curves (total times to successive criteria) for Cond. CC and FC shows that the course of acquisition for these conditions is highly similar. The two curves lie very close to one another and are almost exactly parallel throughout acquisition. (Because of the erratic nature of the distribution of test trials in Cond. CF and FF, learning curves were not plotted for these conditions.)

Total learning times on List 2.—It will be recalled that analysis of performance on the initial test trial of List 1 indicated that the groups assigned to the List 2, FF procedure were of superior learning ability to the groups assigned to the List 2, CC procedure. Inspection of Table 1 further shows that there were large differences in List 1 total learning time between the two FC groups and the two FF groups. Both of these differences are in favor of the List 2, FF condition. In view of what appear to be large inequalities in learning ability between Groups FC-CC and FC-FF and between Groups FF-CC and FF-FF, any findings which indicate that List 2 learning is faster under the FF procedure would have to be viewed with extreme caution. In order to control for the differences in learning ability to some extent, analyses of covariance were used to evaluate transfer-list performance. The correlations between List 1 study time and List 2 total time were generally higher than the correlations between List 1 total time and List 2 total time. Therefore, List 1 study time was used as the predictor variable. Since there were four conditions of first-list learning, there was no common predictor variable, and a separate analysis of covariance was done for each pair of groups assigned to the same List 1 condition.

Table 2 presents the means of the second-list time scores for the eight groups separately, as well as the means for the combined CC and FF groups. It can be seen from Table 2 that there is little overall difference between the CC and FF conditions in study-plus-test time. In none of the four analyses of covariance is the difference between the CC and FF, second-list procedures significant, all Fs $(1, 38) \lesseqgtr 3.36$, $p > .05$.

Since the separate covariance analyses do not allow for any tests of differential transfer as a function of List 1 condition, an analysis of variance was also performed on the unadjusted study-plus-test times of all the groups. This analysis corroborates the results of the covariance analyses in showing that there is no overall difference between the List 2, CC and FF procedures, F (1,

152) < 1. However, the groups who learned List 1 under the self-paced procedures (Cond. FC and FF) performed better on List 2 than the groups who learned List 1 under the *E*-paced procedures (Cond. CC and CF). The mean List 2 total learning time for the groups who learned List 1 under the self-paced procedure was 191.0 sec., while the mean for the groups who learned List 1 under *E*-paced procedures was 239.2 sec. This difference is significant, F (1, 152) = 4.21, $p < .05$. Moreover, it does not appear to reflect any differences in learning ability. On List 1, the total learning times were 356.2 sec. and 358.8 sec. for the groups who learned under self- and *E*-paced conditions, respectively. The interaction of List 1, Self- vs. *E*-Paced × List 2 Condition is not significant, F (1, 152) = 1.91, $p > .05$. Thus, it appears that when *S*s learn their first paired-associate list in the laboratory under a self-paced procedure, they acquire skills which facilitate subsequent paired-associate learning even if the subsequent task is not learned under a self-paced procedure.

TABLE 3

MEAN STUDY TIMES (STs) BETWEEN THE NEXT-TO-LAST AND FINAL TEST TRIALS AND MEAN NUMBERS OF CORRECT RESPONSES (NC) ON THE NEXT-TO-LAST TEST TRIAL

Group	ST	NC
	List 1	
CC	16.0	6.30
FC	28.1	6.30
CF	17.6	5.82
FF	27.6	5.28
	List 2	
CC-CC	16.0	6.45
FC-CC	16.0	6.15
CF-CC	16.0	6.15
FF-CC	16.0	6.05
M Cond. CC	16.0	6.20
CC-FF	21.0	5.85
FC-FF	39.3	5.05
CF-FF	21.9	6.20
FF-FF	32.0	5.25
M Cond. FF	28.6	5.59

Note.—STs are given in seconds.

Evaluation of possible biases in total-learning-time scores.—Since *S*s who learned under Cond. FC, CF, and FF controlled the amount of study time between test trials (either by controlling the duration of the study trials, the number of study trials intervening between test trials, or both), it is possible that *S*s in these conditions overshot the criterion of learning. To the extent that such overlearning occurred, there is a bias against showing that Cond. FC, CF, and FF were superior to Cond. CC, in which the study time between successive tests was always 16 sec. While there is no fully adequate method for evaluating the extent of the bias, an attempt was made to obtain some indication of its size by the following technique: For each *S* in Cond. FC, CF, and FF, the amount of study time intervening between the next-to-last and final test trials was computed. The means of these scores are presented in Table 3. A mean score larger than 16 sec. indicates that there may be a bias against showing a difference from Cond. CC. For List 1, the largest mean study time between the next-to-last and the final test trials is 28.1 sec.; for List 2, it is 39.3 sec. Also presented in Table 3 are the mean numbers of correct responses on the next-to-last test trial. These latter means need to be considered in conjunction with the time scores, because to the extent that *S*s in Cond. FC, CF, and FF showed lower performance than *S*s in Cond. CC on the next-to-last test trial, they would be expected to need more study time than *S*s in Cond. CC to reach criterion. It can be seen in Table 3 that with the exception of Cond. FC during first-list learning and Group CF-FF during second-list learning, the mean number correct on the next-to-last test trial is higher for Cond. CC than for the other conditions. Thus, it appears that if any bias exists against showing that Cond. FC, CF, and FF are superior to Cond. CC, this bias is quite small.

Study time vs. test time.—The amounts of the total time devoted to studying and to testing are shown in Tables 1 and 2. In Cond. CC, study time was necessarily equal to test time, but in Cond. FC, CF, and FF, the proportions of the total time allotted to studying and to testing depended on *S*'s behavior. In these conditions, *S*s spent more time on studying than on being tested. Separate analyses of the study and of the test times show that the four conditions of first-list learning differed in study time, F $(3, 156) = 2.92$, $p < .05$, and in test time, F $(3, 156) = 5.70$, $p < .01$. Similarly, the CC and FF conditions of List 2 differed in study time, F $(1, 158) = 5.31$, $p < .05$, and in test time, F $(1, 158) = 14.70$, $p < .01$. Thus, for both lists, total time is the same, but study and test time differ among conditions. This result is similar to one obtained by Tulving (1967). In a free recall experiment, he varied the relative proportions of study and test trials and found that performance on test trials immediately following study trials was not affected by the distribution of time between studying and recalling. It appears that in some circumstances practice at recalling items may be as beneficial to performance as studying. This would follow if, as Tulving (1967; Tulving & Pearlstone, 1966) has suggested, recall depends on the accessibility of an item as well as its availability in memory storage. Test trials might well function to increase the accessibility of items.

Overt errors.—Little needs to be said about overt errors. All overt errors were expressed as percentages of emissions (correct responses plus errors). Analyses of variance on the error scores support the analyses of total learning times in showing that there are no reliable differences either among the four conditions of first-list learning or between the two conditions of transfer-list learning.

Distribution of study time.—One of the major purposes of the present research was to find out how *S*s distribute study time over the items in a list. Two kinds of analyses of the study durations of *S*s in Cond. FC and FF support the expectation that self-paced *S*s would study difficult pairs longer than easy ones.

For the first kind of analysis, the pairs were ranked 1–8 according to *S*'s overall performance on each pair. A rank of 1 was assigned to the pair *S* got right most often and a rank of 8 to the pair *S* got right least often. The total study times allotted to each pair were also ranked 1–8, with a rank of 1 being assigned to the pair which *S* studied for the least amount of time and rank of 8 to the pair which *S* studied for the most amount of time. Spearman rank order correlations between the difficulty and the study time ranks were computed separately for each *S* in the self-paced conditions. A positive correlation indicates that *S* allotted more study time to the pairs with which he had trouble than to the pairs he found easy. The median correlations were .70, .62, and .74 for Cond. FC and FF of List 1 and Cond. FF of List 2, respectively. Seventy-seven of the 80 List 1 correlations and 78 of the 80 List 2 correlations were positive. The probability of so many correlations being positive by chance is less than 10^{-6}. Also, 44 first-list and 50 second-list correlations were significant at the .05 level. Thus, the correlations provide very strong evidence that *S*s in self-paced procedures study difficult pairs longer than easy ones.

In the second kind of analysis of the study durations, study intervals following an error on the previous test trial were compared with those following a correct response. This analysis was limited to first-list learning under Cond. FC because sometimes no test trials intervened between successive study trials in Cond. FF. The mean study duration following errors was twice as long as that following correct responses (7.12 sec. vs. 3.54 sec.). All 40 *S*s had a mean study duration following errors that was larger than that following correct responses.

Postexperimental inquiry.—The *S*s who learned under *E*-paced procedures were asked whether they had rehearsed all items for equal amounts of time. Of the 80 *S*s who learned List 1 under Cond. CC and CF, 65 reported some tendency to rehearse difficult pairs more than easy ones. For List 2, 68 of the 80 *S*s who learned under the CC procedure made a similar report. The reports included statements of concentrating harder on difficult pairs and of rehearsing difficult pairs during the ITIs and/or during the presentation of pairs learned early. Thus, the reports given by *S*s who learned under the *E*-paced procedures indicate that even when each item is presented for the same amount of time, the covert distribution of study time is related to pair difficulty and to previous performance.

DISCUSSION

In their article on TTI, Cooper and Pantle (1967) distinguished between the "nominal" time (the time potentially available for rehearsal) and the "effective" time (the time during which rehearsal actually occurs). They stated that TTI will occur only when the total effective time remains constant across conditions. This distinction helped Cooper and Pantle delimit some of the boundary conditions for the occurrence of TTI, but it does not deal directly with the essential nature of TTI, which is that different *distributions* of a constant amount of effective study time produce equal performance. For instance, whenever presentation rate is varied and TTI is obtained, it is necessarily the case that the number of individual presentations into which the total time is divided differs among conditions. Why is it that such differences in the way in which study time is nominally distributed do not affect performance? Results of the present study help provide a tentative answer to this question.

The major findings of the present experiment are: (*a*) TTI holds for conditions of practice in which *S* has considerable control over the temporal distribution of his activities, as well as for standard learning procedures; and (*b*) in self-paced conditions, *S*s distribute their rehearsal time unevenly over the items in the list in a manner related to their relative difficulty. Taken together, these two findings appear to yield a somewhat enigmatic conclusion, viz., that *S*s who rehearse at their own rate and who study difficult pairs longer than easy ones do not learn faster than *S*s who are given only 2 sec. on each trial to study each pair. However, the postexperimental reports of *S*s in the *E*-paced conditions indicate that this difference between the way rehearsal time is distributed in *E*- and self-paced conditions is more apparent than real. The majority of *S*s in Cond. CC and CF stated that they covertly rehearsed difficult pairs longer than easy ones and that they tended not to switch their attention from one pair to the next every 2 sec. That is, *S*s in these conditions appear to have covertly distributed their study time over the pairs in a manner similar to the way *S*s in the self-paced conditions overtly distributed their time. It can thus be argued that *E*- and self-paced conditions did not differ in total learning time because the covert rehearsal pattern in the *E*-paced conditions was similar to the overt rehearsal pattern in the self-paced conditions.

In addition to the distinction offered by Cooper and Pantle (1967), a general distinction between the "nominal" and the "covert" *distribution* of study time is called for. The way *S* covertly distributes his study time probably rarely corresponds directly to *E*'s nominal distribution. Assuming this distinction to be valid, an account of the occurrence of TTI in other experiments, such as those that have varied presentation rate, can be derived. Basically, the notion is that the way *S*s covertly distribute their study time is much the same, regardless of the experimental procedure. With regard to variation of presentation rate, it seems likely that when a fast rate is used, *S*s may concentrate on a few pairs on each study trial, and when a slow rate is used, they may rehearse items other than the one which is currently being shown.

REFERENCES

Bugelski, B. R. Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology,* 1962, **63**, 409–412.

Bugelski, B. R., & Rickwood, J. Presentation time, total time, and mediation in paired-associate learning: Self-pacing. *Journal of Experimental Psychology,* 1963, **65**, 616–617.

Cooper, E. H., & Pantle, A. J. The total-time hypothesis in verbal learning. *Psychological Bulletin,* 1967, **68**, 221–234.

Fischer, G. J. Relation between stimulus presentation time, serial learning, and the serial position effect. *Journal of Experimental Psychology,* 1966, **72**, 153–154.

Morsund, J. P., & Chape, C. M. Comparison of three presentation methods in paired-associate learning. *Psychonomic Science,* 1966, **6**, 167–168.

Tulving, E. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 175–184.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 381–391.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

Waugh, N. Free versus serial recall. *Journal of Experimental Psychology,* 1961, **62**, 496–502.

(Received March 27, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 448–452

ABSOLUTE AND RELATIONAL DISCRIMINATIONS INVOLVING THREE STIMULI[1]

MICHAEL D. ZEILER[2] AND ANN G. FRIEDRICHS
Emory University *University of North Carolina*

Children learned to respond to either the largest, the smallest, or the middle-sized member of a stimulus set, and then received transposition tests. Whether relational or absolute stimulus properties controlled responding often depended on the particular relational character of S+. Of the 28 children, 25 transposed in all of the tests following training with a large or a small S+. After training with a middle-sized S+, transposition was less frequent. Twelve children who transposed with a large or a small S+ made absolute choices with a middle-sized S+. Following the transposition tests, *S*s tried to pick S+ from among all of the stimuli used on that day. Whenever *S*s had made absolute choices in the tests, S+ was always identified correctly. However, if *S*s had transposed, they usually could not identify S+ correctly, indicating that they had not learned the absolute properties when they were controlled by the relation.

Transposition experiments provide information about stimulus control during discrimination learning by revealing which aspect of the positive stimulus (S+) controls responding. Little is known about behavior when S+ is one of the extreme members of a set composed of three stimuli. This situation is significant for those interested in analyzing the variables generating control by various stimulus aspects, because the existing data suggest that such conditions may have unique properties. Lashley (1938) reported that rats never learned to respond to the middle-sized stimulus, but learned readily when S+ was either of the other stimuli. Although Lashley ascribed these results to the complexity of the middle-sized as compared with the large or small relation, Spence (1942) deduced the difficulty of the middle-size problem by reference to absolute stimulus properties. Despite Lashley's and Spence's assumptions that *S*s depend on relational or absolute properties in all cases, there is no evidence to support either contention; perhaps the difference in learning speed resulted from a dependence on relations when S+ was the largest or the smallest stimulus and a dependence on absolute properties when S+ was middle-sized. Lashley gave transposition tests after *S*s responded consistently to the large or the small S+ and found control by relations, but the rats' failure to learn when S+ was middle-sized precluded further testing. Thus, although relational control occurred with the large and the small S+, the question of whether the middle-size problem was not learned because it involved a difficult absolute discrimination or a complex relational discrimination remained unsolved.

The ability of children to learn when S+ is middle-sized makes them suitable *S*s for investigating if the controlling stimulus aspect varies when S+ is the largest, smallest, or middle-sized of three stimuli. With respect to middle-size problems, Zeiler and Salten (1967) found that different children learned different aspects of S+. (Comparable data are not available for training with a large or small S+.) The variety in patterns of responding among *S*s suggests that experiments should consider individual data if they are to give a true picture of behavior. In the present research, therefore, each *S* received the three types of training, and consistencies or changes in absolute and relational control were observed for each child.

---

[1] The preparation of this manuscript was supported by Research Grant HD 02845 from the United States Public Health Service. The authors want to express appreciation to Donald Mapel of Meadowbrook Day Camp, Weston, Massachusetts, for his cooperation in providing *S*s and facilities for the accomplishment of this study.

[2] Requests for reprints should be sent to Michael D. Zeiler, Department of Psychology, Emory University, Atlanta, Georgia 30322.

Previous research revealed that the similarity of set members to each other (within-set similarity) determined the frequency of absolute and relational control. Increased within-set similarity increased relational control in a middle-size problem (Zeiler & Gardner, 1966), but decreased relational control when S+ was the larger of two stimuli (Zeiler, 1966). The present study investigated whether the opposite effects of within-set similarity with a large and a middle-sized S+ persisted with the number of stimuli in the sets equated.

METHOD

Subjects.—The 28 children, 17 girls and 11 boys, ranged in age from 5 to 7 yr.

Apparatus.—The stimuli were 16 unpainted .25-in. Masonite square blocks designated in order of increasing size by the numbers 1–16. The surface area of the smallest block, 1, was 1.00 sq. in., and each succeeding block increased in area by a factor of 1.2:1. An indentation in the underside of each square could hide either a plastic chip or a penny.

The squares were displayed on a turntable divided in half by a screen. When turned in one direction, the screen prevented the children from seeing the blocks. When turned in the other direction, the screen was lowered to enable *E* to observe the choices.

Procedure.—The *S*s were assigned at random to three experimental groups with 9 *S*s in each of Groups 1 and 3 and 10 in Group 2. In the course of the experiment, each child learned to respond to the largest, the smallest, and the middle-sized member of a set of three blocks. Group 1 had sets composed of blocks that differed from each other by a factor of 1.2:1, Group 2 by 1.44:1, and Group 3 by 1.73:1. The Group 1 training sets were composed of Stimuli 3-4-5 and 6-7-8; Group 2: 2-4-6 and 8-10-12; Group 3: 1-4-7 and 10-13-16. Each member of the group received both sets. The smaller set was used when S+ was the large stimulus, and the larger set was used when S+ was the small stimulus. Both sets served for middle-sized training, with *S*s being assigned one or the other on a random basis. One of the six possible sequences of the three discriminations was selected on a random basis for each child.

A noncorrection procedure and discrete trials were used throughout the experiments. On each trial, a set was exposed until *S* made a choice. Then, *E* interposed the screen between *S* and the blocks and made any necessary changes. Stimulus positions varied according to a predetermined irregular sequence.

The first day was used to acquaint *S* with the experimental situation and to establish a discrimination maintained by a schedule of intermittent reinforcement. Intermittent reinforcement in training permitted the later introduction of test trials under extinction conditions by minimizing the contrast between training and testing (Zeiler & Salten, 1967).

Day 1 began with continuous reinforcement with the plastic chip for responses to S+ until there were five consecutive correct responses. Whether S+ was the largest, the smallest, or the middle-sized set member varied between children. Then, *E* told *S* that pennies would be hidden instead of the chip, but that *S* would find a penny only after making several correct responses in a row. The *E* emphasized that the same block was always correct and that pennies would only be found occasionally. The children were told that the pennies could either be kept or saved to buy trinkets and the toys that were displayed on shelves in the experimental room. Purchase of toys could be made at the end of each session. The number of correct responses per penny varied from one to six, with the average being three. The first session ended when *S* received 10 pennies.

On Day 2, *E* reminded *S* about the intermittent reinforcement procedure. The training set and S+ were the same as on Day 1. After *S* made six successive correct responses, there was one trial with a test set. From then on, a test trial followed every four to six correct responses with the training set until each of the three test sets was presented four times. No responses were reinforced in the tests. Several trials with the training set followed the twelfth test trial. Finally, *S* was asked to identify the block that had hidden the penny from among the six stimuli that composed the day's training and test sets.

On Day 3, the correct stimulus had a different relation to the members of the training set than it had on Days 1 and 2. The *E* told *S* that the game was the same, but that the pennies would be under a different block. Intermittent reinforcement was used throughout the session. Testing began after six consecutive correct responses and continued as on Day 2. After each test set appeared four times, there were several trials with the training set, and the day ended with *S* being asked to identify S+ when it was presented simultaneously with the other stimuli used in the session.

Day 4 was like Day 3 except that S+ had a different relation to the training set stimuli than it had on any of the earlier days.

Table 1 shows the three sets used with each training set. When S+ was the smallest of three blocks, the test sets were all smaller than the training set; when S+ was the largest block of the training set, the test sets were all larger than the training set. This arrangement guaranteed that the same stimulus was not simultaneously the one closest in size to S+ and the one with the same relation to the other set members as S+ had to the training set. When S+ was the middle-sized block, the test sets could be either larger or smaller than the training set without the same stimulus being both the set member closest in size to S+ and middle-sized. The test sets were selected so that they were one, two, and three steps different from their respective training sets. Within-set similarity was the same in the training and the test sets. With a stimulus series in which each successive member differs equally from the next, successive steps were formed by substituting the next stimulus in the series for the member of the previous set that was the most different from it.

RESULTS

Performance with the training sets.—The number of errors made by individual *S*s in learning the first discrimination to the criterion of five successive choices of S+ ranged 0–7; the number of trials to criterion ranged 5–18. The means and the distributions of errors and trials did not differ for the three groups or when S+ was the largest, smallest, or middle-sized stimulus.

The change from continuous to intermittent reinforcement produced no more than one error. The discriminations on Days 3 and 4 were learned with a maximum of three errors; 26 of the 28 *S*s had either one or no errors. No error ever occurred with a training set after testing began.

Responses in the tests.—The *S*s had four trials with each test set under each of the three training conditions. Figure 1 shows the percentage of *S*s in each group who transposed with all of the test sets. Following training to respond to the largest or the smallest stimulus, all but one *S* in each group always chose the test square having the same relation as S+. Following training to respond to the middle-sized stimulus, the probability of uniform transposition declined as within-set similarity decreased. Figure 1 also shows the percentage of *S*s who made absolute choices, i.e., always responded to the test set member closest in size to S+. One *S* in each group responded absolutely whether S+ was the large, small, or middle-sized block. In addition, four other *S*s in each of Groups 2 and 3 made absolute choices after middle-sized training. The *S*s who neither transposed nor made absolute choices uniformly after middle-size training showed a decreasing frequency of transposition and increasing frequency of absolute choices as the test sets became less similar to the training set.

TABLE 1

TRAINING AND TEST SETS

Group	Training set	Test sets		
		1 step	2 steps	3 steps
1	6-7-8	5-6-7	4-5-6	3-4-5
1	3-4-5	4-5-6	5-6-7	6-7-8
2	8-10-12	6-8-10	4-6-8	2-4-6
2	2-4-6	4-6-8	6-8-10	8-10-12
3	10-13-16	7-10-13	4-7-10	1-4-7
3	1-4-7	4-7-10	7-10-13	10-13-16

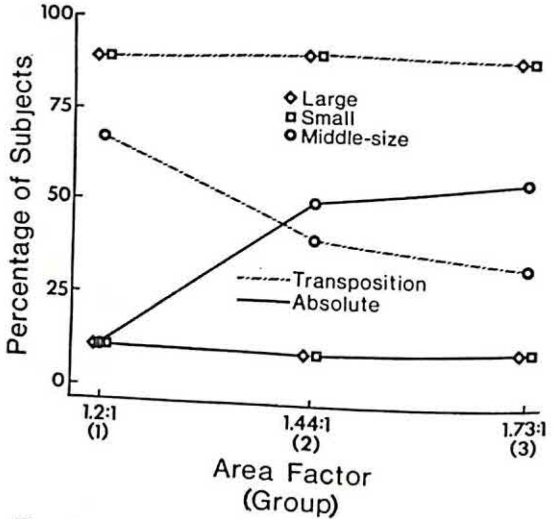


FIG. 1. Percentage of children in each group who transposed with all test sets and who responded absolutely with all test sets.

Because the same *S* was studied with S+ being the large, small, and middle-sized stimulus, it was possible to analyze individual behavior under the different conditions of training. Different patterns of responding were unrelated to either the sex or age of the *S*s. Three Ss made absolute choices, and 13 transposed regardless of the relational nature of S+. The behavior of the remaining *S*s depended on the type of S+. All 12 transposed whenever S+ was either the large or small stimulus. When S+ was the middle-sized block, however, 8 *S*s responded to the block closest in size to S+ in all of the tests. For the other 4 *S*s, there was progressively less transposition and more absolute choices as the steps between the training and the test set increased. Of these 12 *S*s who transposed with a large or small S+ and made either all or some absolute choices following training with the middle-sized S+, 2 were in Group 1, 5 were in Group 2, and 5 were in Group 3. These differences in frequency were statistically significant, χ^2 $(2) = 6.0$, $p < .05$.

At the end of the three sessions involving tests, *S* was asked to select S+ from among the six stimuli comprising the training and the test sets of the day. In the 63 instances of sessions with uniform transposition, S+ was identified correctly 23 times. In contrast, S+ was always identified correctly in the 17 instances when the task followed uniform absolute choices, and in 3 of the 4 sessions when the frequency of transposition and absolute choices depended on the steps between the training and the test set. No *S* was correct on every day on which he or she transposed with all of the test sets. The proportion of identifications was about equal following uniform transposition with either the large, small, or middle-sized stimulus as S+.

DISCUSSION

The choices following training to respond to the middle-sized stimulus replicated and extended Zeiler and Salten's (1967) finding that different children were controlled by different stimulus aspects. Some were controlled by the middle-size relation and others by absolute size. Still others transposed with the test set most similar to the training set and then made an increasing number of absolute choices as the test and training sets became more different from each other. This behavior may reflect an inaccurate discrimination of absolute size, or simultaneous control by absolute and relational properties. The first alternative is the more parsimonious one since the changes in responding are predicted by a model that transforms absolute size into its ratio to the context of stimuli (Zeiler, 1963). Further argument against an explanation based on the learning of both absolute and relational properties is that other research showed that *S*s who learned both aspects of stimuli had the opposite pattern of choices: they responded to the stimulus closest in size to S+ with similar training and test sets and transposed only when the differences between the training and test sets were increased (Zeiler, 1964).

Except for the one *S* in each group who always responded to the stimulus closest in size to S+, all of the *S*s transposed following training to respond to the largest or the smallest training block. Thus, although the present experiment did not replicate Lashley's (1938) finding that *S*s were able to learn a discrimination when S+ was the large or the small stimulus, but not when it was middle-sized, it did replicate Lashley's observation of uniform transposition when S+ was the largest or the smallest set member. Some *S*s also transposed whenever S+ was middle-sized, but the majority in Groups 2 and 3 were controlled by absolute properties following middle-size training. These data replicated Zeiler and Gardner's (1966) finding of a positive relation between transposition and within-set similarity. In general, then, the frequency of relational control depended on the particular relational character of S+ and was influenced by within-set similarity only when S+ was the middle-sized stimulus.

Large and small stimuli do not produce relational control under all conditions. Zeiler (1966) found that *S*s of the same age as those of the present study transposed after being trained to respond to the larger of two squares only when the set was composed of stimuli quite similar to each other. The present finding of transposition following training with a large or a small S+ independent of within-set similarity seems to be due to the use of three rather than two stimuli. Relational control with three stimuli may occur because the large and small stimuli have the same relation to both of the other set members. The frequent shifts to absolute control with a middle-sized S+ are perhaps the product of the middle-sized block having a different relation to each of the other set members.

The results of another study (Zeiler, 1965) suggested that the present findings could not be attributed to the within-*S* design. In that experiment, separate groups of children were trained to respond to the largest, smallest, and middle-sized of three stimuli, and individual *S*s received either one or two test sets. The groups trained with a large or a small S+ showed nearly perfect transposition; the group given middle-size training made many absolute choices.

The posttransposition test data of the present experiment showed that the same *S*s may or may not learn the absolute size of S+ when they are controlled by the relations. The *S*s could select S+ on the basis of its absolute properties in less than 40% of the sessions of uniform transposition. In contrast, whenever *S* always responded to the test stimuli closest in size to S+, and, in most of the other cases in which there was evidence for absolute control, S+ was identified correctly. Thus, the same *S* would accurately select the stimulus on a day of absolute choices and be incorrect on a day of uniform transposition. Stoddard (1968) reported a related phenomenon: children frequently could not respond on the basis of the absolute properties of S+ after it had appeared as the odd member of a multistimulus array. Although the learning of absolute properties may accompany the learning of a relation, it often does not in either size or oddity discriminations. Whether *S*s controlled by absolute properties also learn the relation of S+ to the other members of the training set remains an open question.

REFERENCES

LASHLEY, K. S. The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *Journal of General Psychology*, 1938, **18**, 123–193.

SPENCE, K. W. The basis of solution by chimpanzees of the intermediate size problem. *Journal of Experimental Psychology*, 1942, **31**, 257–271.

STODDARD, L. T. An observation on stimulus control in a tilt discrimination by children. *Journal of the Experimental Analysis of Behavior*, 1968, **11**, 321–324.

ZEILER, M. D. The ratio theory of intermediate size discrimination. *Psychological Review*, 1963, **60**, 516–533.

ZEILER, M. D. Transposition in adults with simultaneous and successive stimulus presentation. *Journal of Experimental Psychology*, 1964, **68**, 103–107.

ZEILER, M. D. Transposition after training to the largest, smallest, or middle-sized of three stimuli. Paper presented at the meeting of the Eastern Psychological Association, Atlantic City, April 1965.

ZEILER, M. D. Solution of the two-stimulus transposition problem by four- and five-year-old children. *Journal of Experimental Psychology*, 1966, **71**, 576–579.

ZEILER, M. D., & GARDNER, A. M. Intermediate size discrimination in seven- and eight-year-old children. *Journal of Experimental Psychology*, 1966, **71**, 203–207.

ZEILER, M. D., & SALTEN, C. S. Individual gradients of transposition and absolute choice. *Journal of Experimental Child Psychology*, 1967, **5**, 172–185.

(Received March 27, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 453–461

WHY ARE RHYMES EASY TO LEARN?[1]

GORDON H. BOWER[2] AND LAURA S. BOLTON

Stanford University

The question was why a list of rhyming paired associates is learned faster than a list of unrelated pairs. It is proposed that the rhyming relation restricts the range of response alternatives to the stimulus, practically converting recall into a recognition test. Several tests of this hypothesis proved confirmatory. First, an assonance (change of last phoneme) rule for pairs, which restricts alternatives about as much as a rhyming rule, facilitated performance as much as did a rhyming rule. Second, when response alternatives were equated by multiple-choice tests for memory, the advantage for rhyming pairs vanished. A third experiment showed that the presence of some rhyming pairs in a list induced *S* to generalize this rule inappropriately to other pairs, thereby suffering interference on those pairs composed by re-pairing rhyming units.

It is a well-known fact of everyday life, and even of the laboratory, that rhyming is an aid to memory. Poetry, at least rhyming verse, is more easily learned than prose (McGeoch & Irion, 1952). The popular mnemonic systems (e.g., Furst, 1948) propound and exploit the power of rhymes, and most of us have learned the alphabet, the number of days in the months, and the order and names of cranial nerves, etc., by one or another rhyming mnemonic. These mnemonics, or the poetry, songs, and limericks which they resemble, display two organizing features for the memorizer: rhythmic pattern, or meter, and rhyming of the end word in successive repetitions of the rhythmic pattern.

The mnemonic function of rhythmic pattern may stem from its counting and place-keeping features, i.e., it tells *S* the number and order of stressed syllabic units in each constituent (line). But putting aside any analysis of rhythmic pattern, we shall here focus on the rhyming feature of poetry. Paring matters down to the merest bones, rhyming is a relationship between two words or phonological compounds. Specifically, a word which rhymes with "hat" is a word obtained by altering the first phoneme but leaving the remaining phonemes the same (as in "cat," "mat," and "bat").

This suggests that one can study the influence of rhyming at the elementary level of paired associates, as was done here. There is little doubt that adult *S*s would learn a list of rhyming pairs much faster than a list of unrelated pairs. The interesting question is why this is so. What is it about rhymes that makes them memorable? Is it their frequency of cooccurrence in the language? Is it common associative linkages? Is it simply a result of formal graphemic similarity? Will rhymes be better in all circumstances, or can one create conditions where they lose their advantage over unrelated pairs? Can one construct a list of nonrhyming pairs which, because of a simple structure, will confer the same memorial benefit as do rhyming lists?

One obviously seeks for some hypothesis to explain why rhymes work in most cases. The search need not go very far because on closer examination, an elementary hypothesis suggests itself. The hypothesis is that rhymes are normally easily learned because knowledge of the rhyming relation considerably restricts the range of response alternatives that need be considered for a given stimulus. That is, knowing the stimulus term and knowing that the correct response rhymes with the stimulus, one is able to exclude many list responses from consideration, restricting consideration to only the rhyming candidates. Furthermore, this knowledge suffices to reduce an ostensible recall task almost to a recognition task since *S* can often implicitly generate most of the plausible rhyming responses and he need only select that response candidate recently experienced in this list context.

---

[1] This research was supported by Research Grant MH-13950, awarded to the first author from the National Institute of Mental Health. Thanks are due to Peter Arnold, who helped with the last experiment, and to Gary Olson, who contributed some original ideas.

[2] Requests for reprints should be sent to Gordon H. Bower, Department of Psychology, Stanford University, Stanford, California 94305.

This hypothesis supposes that there are two important components that determine performance on rhyming pairs: First, *S* must know or believe that a given stimulus has a rhyming response; and second, this knowledge then enables him to restrict the response alternatives to the stimulus. The former knowledge can be given by a whole-list rule, viz., all pairs rhyme, which *S* will quickly learn. Or it can be provided by an external discriminative stimulus, as was done here (viz., stimulus words printed in red have rhyming response terms). But even if these external cues are lacking, *S* may nonetheless adopt a rhyming strategy for generating guesses to unlearned items if there are sufficient rhyming pairs in the list to make this strategy worthwhile. So even without cues distinguishing rhyming from unrelated pairs, performance may be superior on rhyming pairs due to such strategic response biases.

The three experiments to be reported were testing various aspects of this hypothesis to explain the memorial benefits of rhyming. In the first two experiments, by instructions plus a discriminative stimulus, *S* knew which stimuli in the list had a rhyming response and which had an unrelated response. In the last experiment, this external cue was lacking, and *S* had to figure his way through a very interfering list of scrambled rhyming words.

EXPERIMENT I

The first experiment was designed for three aims, and for each a particular group of *S*s was run. The first aim was simply to demonstrate, for the formal record, that when appropriate discriminative information is available to *S*, he will perform much better on rhyming pairs than on unrelated word pairs. The second aim was to demonstrate that a facilitation similar to that conferred by rhyming could be produced by a different relational rule that does not involve rhyming. The relationship between the cue and response term in this case was a simple change in the *last* phoneme (e.g., HAT–*ham*, BIN–*bit*). With CVC word pairs, this assonance rule appears to be about as restrictive of response alternatives as is the rhyming rule. That is, for the CVC stimuli used, approximately the same average number of CVC response words could be derived from it using either the rhyming or the assonance rule. If the assonance rule is approximately as restricting of response alternatives as is the rhyming rule, it should produce a similar magnitude of facilitation in performance as compared to unrelated word pairs.

The third aim was to show more directly that the restriction of response alternatives was the main effect of the rhyming relation. To this end, a multiple-choice recognition test for memory was used in which the rhyming pairs and unrelated word pairs were equated with respect to restriction on the number of response alternatives. Furthermore, in half of the cases, the rhyming information was rendered useless by having all the multiple-choice alternatives rhyme with the stimulus term (e.g., the cue HAT and response alternatives *fat, cat, mat, bat, pat*). The idea was that by rendering the rhyming rule nondiscriminative with respect to these response alternatives, the performance difference between rhyming and unrelated pairs would disappear completely.

Method

Subjects.—The *S*s were 30 Stanford undergraduates fulfilling a service requirement for their introductory psychology course. They were run individually and assigned in random order to the three experimental conditions, with equation of male–female proportions in each group. There were 9 *S*s in the rhyme and end-change groups and 12 in the multiple-choice group.

Design and procedure.—The three groups each learned one list of 36 paired associates for nine anticipation trials; the pairs were tested and studied at a 2:2-sec. rate on a memory drum and shown in a new random order on each trial. All stimulus and response terms for all lists were CVC words, almost all nouns. The list for Group R (rhymers) consisted of 18 rhyming pairs typed in red letters mixed in with 18 unrelated pairs typed in black letters. The list for Group E (end-change) consisted of the same 18 unrelated pairs as in Group R plus 18 pairs in which the response word was related to the stimulus word by altering the last letter (phoneme). The latter 18 stimulus terms were the same as in the R list; only the response terms differed. In List E, the end-change pairs were typed in red and the unrelated pairs in black. The list for Group M (multiple-choice) consisted of 36 pairs divided into 6 pairs each of six different types, as follows: (*a*) rhyme–rhyme (RR) pairs, in which the stimulus and response rhymed and all four distractors on the multiple-choice tests rhymed with the correct response; (*b*) rhyme–unrelated (RU), in which the stimulus and response rhymed, but the four distractors were unrelated to the correct response; (*c*) unrelated–rhyme (UR), in which the stimulus and response were unrelated, but the four distractors rhymed with the correct response; (*d*) unrelated–unrelated (UU), in which the stimulus and response were unrelated and the four distractors were unrelated to the correct response; (*e*) rhyme (R), in which the stimulus and response rhymed and *S* had to recall or anticipate the correct response, not select it from a multiple-choice list; and (*f*) unrelated (U), in which the stimulus and response were unrelated and *S* had to recall or anticipate the correct response.

In total, there were 18 rhyming pairs and 18 unrelated pairs, 24 pairs tested by multiple-choice recognition and 12 pairs by anticipation recall. All pair types appeared in a new scrambled order on each trial. The cue term for the rhyming pairs was typed in red, and the stimulus–response pair during the study interval was also in red. The unrelated pairs were typed in black letters. On recognition frames, the cue term appeared to the left, followed by a dash and then five response alternatives, the correct one plus four distractors in random order. After 2 sec., the memory drum turned to display the correct cue–response pair for 2 sec. For a given recognition item, the same five multiple-choice alternatives appeared over all nine trials, but in a different left–right order on each trial. The R vs. U recall items in this third list are essentially a small replication of the comparison provided in List R.

Before the experiment began, all *S*s were informed about the nature of their list, in particular that red cue terms had rhyming response terms (or end changes in List E). They were instructed to use this knowledge to make a guess even on the first trial before they had seen the correct S-R pairing.

Results

Recall learning.—Group R provides a within-list comparison of the ease of learning rhyming vs. unrelated pairs, and Group E provides a similar comparison for end-change pairs. The mean proportions of correct responses over trials for the two subsets of items in each list are shown in Fig. 1, Group R in Panel a and Group E in Panel b.

In Group R, there was a large advantage for rhyme pairs: they started at a higher level on Trial 1 and maintained their superiority over all nine trials. The mean correct responses over the nine trials per item per *S* were 5.95 for rhyme pairs and 4.11 for unrelated pairs. The mean difference, 1.84, is highly significant, $t\ (8) = 4.26$, $p < .001$.

In Group E, there was a similarly large advantage for the end-change pairs over the

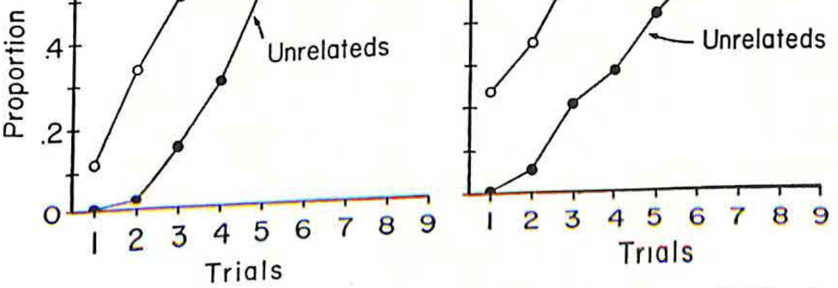


FIG. 1. Learning curves for a mixed list of (*a*) rhyming vs. unrelated pairs or (*b*) assonant (end-change) vs. unrelated pairs.

unrelated pairs: they started at a higher level on Trial 1 and maintained their superiority over all nine trials, with some irregularity in learning after Trial 6. The mean correct responses over the nine trials per item per *S* were 5.38 for end-change pairs and 3.61 for the unrelated pairs. The mean difference, 1.77, is highly significant, $t\ (8) = 6.19$, $p < .001$.

The end-change pairs were somewhat more difficult than the rhyming pairs, but their means do not differ reliably. Similarly, the unrelated pairs are slightly, though not significantly, harder in the E list context than in the R list context. Importantly, the advantage conferred by the end-change rule (1.77) is not reliably different from the advantage conferred by the rhyming rule (1.84), $t\ (16) = .21$, $p > .20$. This indicates that the rhyming effect can be mimicked by employing nonrhyming rules that are approximately equally restricting of the number of response alternatives. Thus, there is nothing uniquely special about the rhyming relation.

The R and U recall items in List M replicated the main results shown in Fig. 1a for List R. The R pairs were consistently ahead of U pairs from Trial 1 onward. The mean correct responses over nine trials per item per *S* were 4.97 for R pairs and 3.89 for U pairs in List M.

In comparing performance on the R vs. U items (or E vs. U items), it should be noted that both the stimulus and response words differed in the two sublists. Stimulus and response factors could have been equated in a between-*S* design, with some *S*s learning only the R pairs and other *S*s learning the R pairs scrambled. But the latter condition might create associative interference, with the two rhyming words in the scrambled R list competing in recall to an unrelated stimulus, and that effect could contaminate the between-*S* comparison of R and U lists. Since the words used were familiar CVC nouns and stimuli were randomly selected for the large R and U item pools, it was believed that the R and U sets would be approximately equated in terms of frequency, meaningfulness, and concreteness of the items. In any event, it was felt that any differential in these factors remaining after random selection of items would have small effects on learning relative to the predicted effect of the pairwise rhyming rule.

Since Fig. 1a and 1b show that R or E pairs start ahead of U items on Trial 1, the question naturally arises whether the advantage for these pairs is merely in initial "guessing" probabilities, with no fundamental difference in learning rate once the differential guessing is partialed out of the performance (cf. Bower, 1962). To answer this question requires a presupposition about the learning curve and its parameters. It shall be answered with respect to the assumption that the mean learning curve for a given type of item can be described by the equation

$$p_n = 1 - (1 - p_1)(1 - \theta)^{n-1}, \qquad [1]$$

where p_n is the proportion correct on Trial n, p_1 is the proportion correct on Trial 1, and θ is the learning-rate parameter. This is a growth curve derivable from a variety of underlying assumptions about how learning occurs. Since *S*s were allowed to omit responses, the parameter p_1 summarizes the tendency to respond times the probability of guessing correctly given a response. Since R and E items were colored red and *S* was instructed regarding the significance of this cue, the tendency to respond was higher for R and E items than for U items. Sheer guesses on U items could hardly ever be correct. The analysis that follows would not be materially changed if p_1 is interpreted as simply the initial tendency to give the correct response under the given experimental conditions.

In terms of the parameters in Equation 1, we can be reasonably confident that the rhyming or assonance rules increase p_1 above that for unrelated pairs. The question is whether these rules also produce a difference in θ, the learning-rate parameter. To answer this question, θ values were estimated for each *S* for his two subsets of items (R vs. U or E vs. U) by (*a*) estimating p_1 from *S*'s proportion correct for each subset on his Trial 1 guess and (*b*) equating the sum of p_n from Trials 1–9 to his observed mean correct responses per item in that subset over the nine trials. The resulting θ estimates should be approximately normally distributed, and the difference between the two θ values for a given *S* (R-U or E-U) should also be normally distributed.

For Group E, the mean p_1 estimate was .250 for end-change items and .006 for unrelated items. The average θ value was .17 for E items and .14 for U items, and these are not significantly different. For Group R, the mean p_1 estimate was .123 for rhyme pairs and .006 for unrelated pairs. The average θ value was .28 for rhyme pairs and .17 for unrelated pairs. These latter θ values differ reliably by a paired *t* test, $t\ (8) = 2.63$, $p < .01$. Examining the R vs. U learning curves in Fig. 1a, however, the differential learning rates for R and U items are apparent only over the first trial or two. For example, the increment in correct response probability from Trial 1 to Trial 2 was 22% for R items and only 2% for U items. That the differential in net θ values is caused by the first-trial increment was shown in a reanalysis using Equation 1, but altered by replacing p_1 by the observed p_2 and estimating θ over Trials 2–9. By this analysis, the θ values for R and U items no longer differ reliably, $t\ (8) = 1.19$. This result suggest that on Trial 1, *S*s predominately selecte the rhyming pairs for rehearsal and learning postponing learning of the unrelated pair for later trials.

With the latter emendation, these analyse of learning rates are consistent with the basi hypothesis; i.e., S-R rules restrict the range of response alternatives so that p_1 is strongly affected, while the learning rate θ is no significantly affected.

Recognition learning.—Learning curves for the four types of recognition items are shown in Fig. 2; the first letter (R or U) in the designation of the item type refers to the nature of the correct S-R pair, and the second letter refers to the relation between the correct response and the distractors on the multiple-choice test. The RU items begin at a high level on Trial 1 and remain high. This was expected, of course, since knowledge that red items rhyme is sufficient to enable *S* to select the correct response without further training on these RU items. The only puzzle is why this curve is not at 100% on all trials. It is suspected that disruption of perfect performance was caused

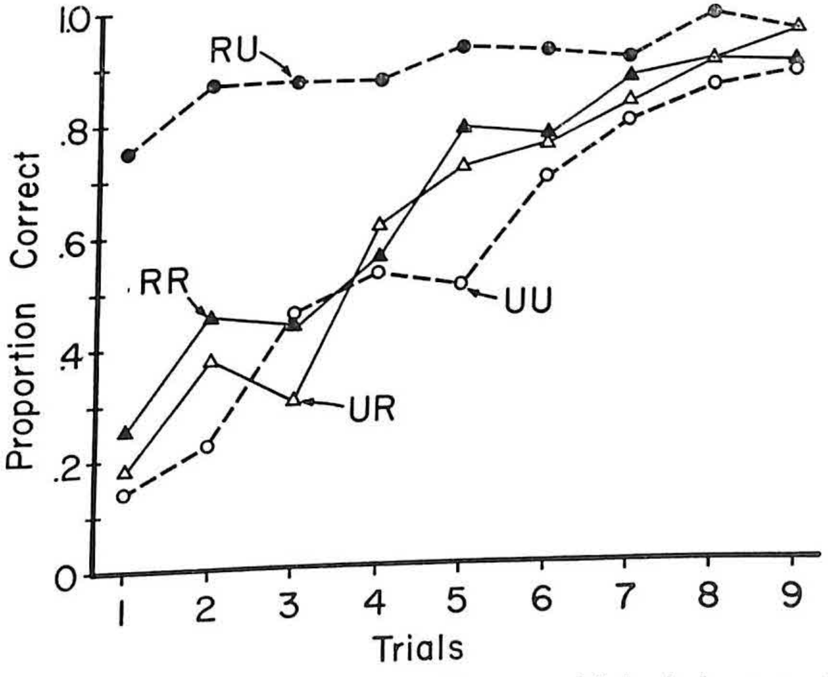


FIG. 2. Proportions of correct S-R recognitions over trials for the four types of recognition items in List M.

by the brief 2-sec. test interval, during which *S* had to read the stimulus word, classify it as red and a rhymer, and then look for a rhyming response among five alternatives.

The remaining three types of items, RR, UR, and UU, have similar learning curves, with no type being clearly superior to the others. They begin on Trial 1 at the chance level of 20%, as they should. An overall measure of total correct responses in nine trials per item per *S* gave 6.04 for RR items, 5.65 for UR items, and 5.04 for UU items. An overall *F* test for repeated measures on these three means yielded significance at the .05 level, $F\ (2, 22) = 4.74$.

The recognition procedure reduced the previous difference between rhyme and unrelated pairs found with the recall procedure in List R. However, the hypothesis had predicted that these three types of recognition pairs, RR, UR, and UU, would not differ since they were equated on the number of response alternatives and since, for RR items, the rhyme rule was rendered useless for selecting the correct response. Therefore, the significant difference in this three-way comparison was disturbing to the authors, and this led to Exp. II.

EXPERIMENT II

According to the hypothesis, the RR, UR, and UU recognition items in List M should not have differed in difficulty, yet they did at a marginal level. Why should that be? One possibility is that the RR and UR items were easier than the UU items because in the former cases, *S* really had to learn only the first letter of the correct response since the last two letters of all alternatives in RR and UR items were the same. This seemed implausible, given the usual belief that words are treated as intact units. A second possibility is that the three types of items were represented by an unfortunate specific sample of pairs that, by chance, turned out to be unequally difficult for idiosyncratic reasons unrelated to the basic hypothesis under test. By the nature of the experiment, one could not counterbalance over *S*s the condition to which a particular pair was assigned. What one can do, however, is to try to replicate the result with a larger set of new pairs of the three types. If the RR, UR, and UU differences are real and important, then they should appear again with a new and larger sample of these item types. If, on the contrary, the former differences resulted from an unfortunate sample of items, the differences due to item types should vanish upon replication with new and larger samples.

Method

Design and procedure.—A single group of 12 *S*s (Stanford undergraduates) had nine trials on a 36-item recognition list consisting of 12 items each of Types RR, UR, and UU. The RR items were typed in red, and the UR and UU items were typed in black. The *S*s were instructed about the significance of this cue before the experiment started.

The 12 instances of each item type were new, with some derived partially from items used in List M of Exp. I. For example, six of the UR items were composed by assigning the former RR response terms to a former UU stimulus form. Six of the new RR items were composed by adding new R terms to the stimuli from the former UR pairs. Six of the new UU items were made by assigning a former UU set of response alternatives to a former RR stimulus term. The remaining six items in each condition were composed, some anew and some from the 18 R and U items in the former List R of Exp. I, using only those R and U items that had been of medium difficulty. It was hoped that by these complex machinations, we had equated or randomized out any differences in idiosyncratic difficulty of the items assigned to the three types.

The 36 items were presented in a new scrambled order each trial, with a guess required on Trial 1. The anticipation interval was lengthened to 4 sec. to give *S* more time to select and say aloud his response. After the anticipation interval, the correct S-R pair was shown for 2 sec.

Results

The mean learning curves for the three types of items are shown in Fig. 3. They start together at 20% on Trial 1 and are closely intertwined throughout the course of learning. There are no apparent differences in learning among the three item types, and this is confirmed by further analyses. The mean total numbers of correct responses over nine trials per item per *S* were 6.07 for RR items, 6.18 for UR items, and 6.13 for UU items. These means obviously do not differ significantly.

The lack of differences here with this larger sample of items makes plausible the presumption that the prior differences in List M of Exp. I resulted from misfortunes of small samples. If so, then the lack of differences in Exp. II would support the hypothesis, viz., that rhyming pairs benefit from restriction of response alternatives and that the effect vanishes when the alternatives are similarly restricted for nonrhyming pairs. However, since the anticipation interval was shorter in Exp. I (2 sec.) than in Exp. II (4 sec.), it is possible that the rhyming effect, that RR items exceed UU items in recognition, is genuine (not due to chance selection of items) but rate sensitive. That is, in order to perform successfully with brief anticipation intervals, *S* might be forced into recalling from the cue alone because there may be insufficient time for him to read and classify the stimulus word, then read five response alternatives, recognize the correct one, and say it aloud. If rapidly paced recognition tests encourage *S* to use the recall mode, then rhyming pairs would be favored over unrelated pairs, as Exp. I (see Fig. 1a) has shown. This interpretation requires checking by running the recognition list of Exp. II with the 2-sec. anticipation interval.

EXPERIMENT III

Experiment I showed that pairs exemplifying a rhyming relation are facilitated in performance relative to unrelated pairs. Experiment III was designed to show that inappropriate generalization of a rhyming rule could *retard* learning of selected pairs in the list. The 24-item list in Exp. III consisted of eight rhyming (R) pairs, eight unrelated (U) pairs, and eight rhyming-interference (RI) pairs, all printed in black letters, so there was no cue informing *S* which were the rhyming pairs. The RI pairs were constructed by re-pairing the stimulus and response words from eight rhyming pairs. For example, from the normal rhymes PIN–*fin*, CAT–*hat*, TAN–*ban*, the RI pairs might be

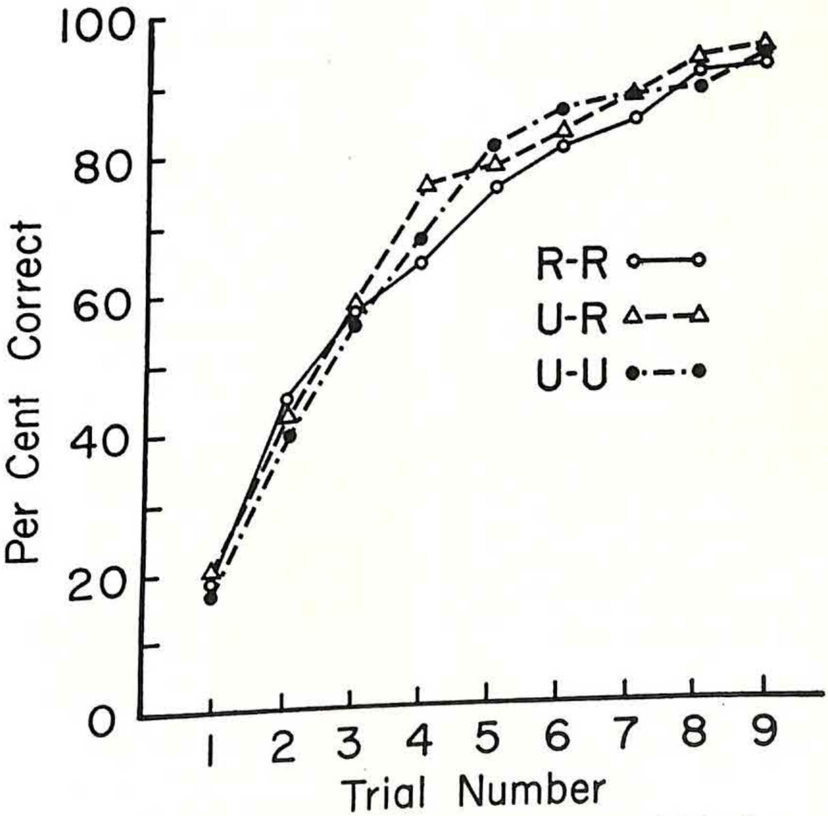


FIG. 3. Proportions of correct S-R recognitions over trials for the RR, UR, and UU items of Exp. II.

PIN–*hat*, CAT–*ban*, and TAN–*fin*. The difference between such RI pairs and the U pairs is that for the RI pairs, the rhyming unit to the stimulus is an available response in the list, but is assigned to another stimulus; on the other hand, the U stimuli did not have a specific rhyming response primed into availability by the list.

The *S*s were informed that some of the pairs in the list conformed to a rhyming rule. Since there were no cues distinguishing the rhyming items from the others, it was predicted that *S* would generalize this rhyming rule to other stimuli, using it as a response-selection strategy for unlearned pairs. Such inappropriate rule generalization should have particularly debilitating effects on the RI items, producing many intrusions and much response competition from their rhyming (incorrect) responses available in the list.

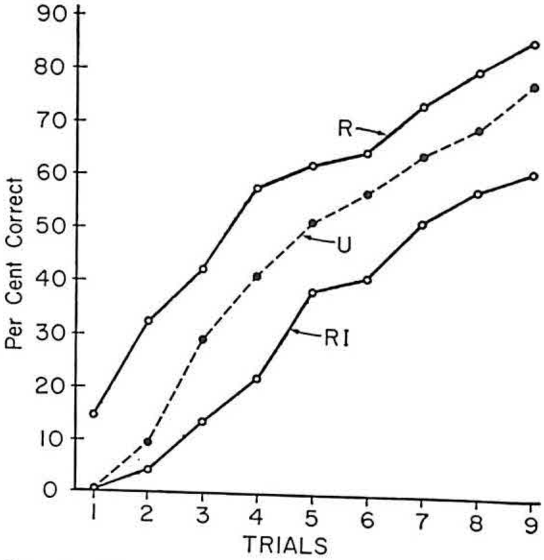


FIG. 4. Mean correct recall percentages for the R, U, and RI pairs of Exp. III.

Method

Each *S* learned a 24-pair list by the anticipation method for nine trials, with a guess requested on Trial 1. The anticipation interval was 4 sec., and the S-R study interval was 2 sec., both controlled by external timers to a memory drum. The 24 pairs were selected from the lists of Exp. I. The 8 rhyming pairs were those of medium difficulty from Cond. R of Exp. I; the 8 U pairs were those of medium difficulty from the U pairs of Exp. I. The 8 RI pairs were composed by randomly re-pairing 8 remaining pairs from the R condition of Exp. I. The 24 items were printed in black type and occurred in a scrambled order each trial. The *S*s were told that the list to be learned contained a number of rhyming pairs and that they should try to guess the correct response even on the first trial.

The *S*s were nine males and nine females attending summer classes at Stanford. Eight *S*s were fulfilling a service requirement for their introductory psychology course, whereas 10 *S*s were paid for their participation. This variable produced no difference in learning proficiency, so the two subgroups were pooled. Each pair of *S*s received a different random order of the 24 items within each trial and varying across trials.

Results

The performance curves for the different types of items are shown in Fig. 4. The conditions are consistently ordered, with R items best and RI items worst. The mean numbers of correct responses over nine trials per item per *S* were 5.12 for R items, 3.97 for U items, and 2.91 for RI items. The overall *F* test comparing these means is significant ($df = 2$, 51, $p < .001$), and Newman-Keuls comparisons show that all pairwise differences are significant at least beyond the .025 level. So, relative to the unrelated pairs, rhyming facilitates and rhyming interference retards performance.

Further analyses revealed relevant sources of errors. Next to omissions, the largest class of errors are intruding responses that rhyme with the stimulus word. For R and U items, these rhyming errors are necessarily extralist intrusions; they account for 58% and 27%, respectively, of the overt errors to R and U items. There are many more rhyming errors to RI items (2.5 times the number for U items); 44% of these intrusions to an RI item came from its lone rhyming response in the list (which was not paired with the RI stimulus). A rhyming response primed into availability by its appearance in the list thus intrudes inordinately, compared to any particular extralist rhyming response.

To summarize the results of Exp. III, the presence of some rhyming pairs in the list leads *S* to adopt a rhyming response bias on unlearned pairs, especially if the rhyming response is available in the list. The consequence of this generalized rule is facilitation of performance on R pairs and interference with performance on RI pairs. In a small experiment run to secure this analysis, nine *S*s were run 10 trials on a list of 24 pairs composed only of U and RI items in equal proportions, and the instructions contained no mention of rhymes. Since *S*s should not adopt a rhyming strategy in this case, one would expect no difference in performance here to the U and RI items. This was verified: mean correct responses over the 10 trials per item per *S* were 4.77 for U items and 5.13 for RI items. The small difference (in the opposite direction) is not significant, $t\ (8) = 1.15$, $p > .10$. This pattern of results means that the differential error rates for U and RI items depend on *S* having a rhyming response bias, which was induced in Exp. III by instructions and the presence of rhyming pairs in the list.

DISCUSSION

A final comment is required concerning the analysis of how S-R rules facilitate performance. It appears likely that a response-restricting rule will facilitate performance only if *S* can in fact readily generate elements of the restricted response set from the stimulus plus rule. The rhyming and assonance rules, which facilitated performance in Exp. I, restrict the response term so as to match the stimulus term in the final or initial phonemes, respectively. These are good cues for generating alternatives. Experiments by Horowitz, White, and Atwood (1968) have shown that initial and final phonemes are good cues for retrieving words from memory, whereas middle fragments of a word are appreciably poorer cues. For instance, __ogn__ is a poor retrieval cue for the word "recognize." Such results tell us something about the phonetic indexing of words in long-term memory.

Reflections on the phonetic organization of the verbal store suggest an alternative formulation of the present hypothesis about rules. The alternative formulation is that the stimulus word plus the rule specify an indexing code with which *S* can enter his memory to locate the response word. This index either directs *S*'s search to a restricted region of memory or, alternatively, is used like a "call number" to retrieve a restricted set of responses from which the list response can be selected (because it has been recently tagged in the experimental context). But if long-term memory for words is organized phonemically in specific ways, then certain types of indexing codes will be relatively useless in that they will retrieve nothing. A literal interpretation of the Horowitz et al. (1968) results is that words in long-term memory are not filed or indexed according to their middle phonemes. This would imply, e.g., no facilitation of performance if the stimulus and response words of a pair overlapped, say, in the middle three of nine letters. Most *S*s would simply be unable to find or generate acceptable candidate responses from such an index code.

REFERENCES

BOWER, G. H. An association model for response and training variables in paired-associate learning. *Psychological Review*, 1962, **69**, 34–53.

FURST, B. *Stop forgetting.* Garden City, N. Y.: Garden City Books, 1948.

HOROWITZ, L. M., WHITE, M. A., & ATWOOD, D. W. Word fragments as aids to recall: The organization of a word. *Journal of Experimental Psychology,* 1968, **76**, 219–226.

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: McKay, 1952.

(Received March 29, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 462–466

EFFECTS OF FACILITORY AND INHIBITORY SETS ON GSR CONDITIONING AND EXTINCTION[1]

MICHAEL E. DAWSON[2] AND PAUL REARDON[3]

California State College, Los Angeles

The effects of facilitory ("It is intelligent to become conditioned,") and inhibitory ("It is intelligent to not become conditioned,") set on the classical conditioning of the GSR of 80 college students were investigated. Both acquisition and extinction data revealed: (*a*) GSR magnitude of the facilitory group was significantly larger than that of the inhibitory group ($p < .05$); (*b*) GSR magnitude of the facilitory and inhibitory groups was significantly larger than that of a pseudoconditioning group ($p < .01$); and (*c*) GSR magnitude of neither the facilitory nor the inhibitory group differed significantly from that of a neutral instructions conditioning group. Differential UCR habituation was associated with the differential conditioning of the facilitory and inhibitory groups. These data suggest that previously reported significant results were due to the summation of facilitory and inhibitory set effects and not to their individual effects.

It has been a common finding that instructional variables affect the course of human classical conditioning. As Grings and Lockhart (1963) pointed out, the instructional variables have been usually of two general types: (*a*) those which inform *S* about how to respond (Hilgard & Humphreys, 1938; Nicholls & Kimble, 1964; Norris & Grant, 1948) and (*b*) those which inform *S* about stimulus relationships (Bridger & Mandel, 1964; Cook & Harris, 1937; Dawson & Grings, 1968).

Until recently, the first type of instructional variable has been studied only with conditioning of skeletal responses. However, Hill (1967) recently employed a variation of the first type of instruction with conditioning of the GSR. Hill indirectly informed *S*s how to respond by instructing a facilitory group that it was "the most adaptive, sensible, and intelligent thing *to become conditioned*" and an inhibitory group that it was "the most adaptive, sensible, and intelligent thing *not to become conditioned.*" The results revealed that the magnitude of the GSR during conditioning was significantly larger for the facilitory group than for the inhibitory.

Unfortunately, all that one may infer from Hill's (1967) data is that the facilitory group responded more than the inhibitory group, which, although interesting, does not answer two primary questions: (*a*) Were the inhibitory instructions successful in completely inhibiting GSR classical conditioning? and (*b*) Were the facilitory instructions successful in significantly facilitating GSR classical conditioning? In order to answer the first question, it is necessary to compare the response magnitude of the inhibitory group with that of a pseudoconditioning group. If the inhibitory instructions were successful in completely inhibiting conditioning, then the response magnitude of the inhibitory group should not differ from that of the pseudoconditioning group. In order to answer the second question, it is necessary to compare the response magnitude of the facilitory group with that of a neutral instructions conditioning group. If the facilitory instructions were successful in significantly facilitating conditioning, then the response magnitude of the facilitory group should be significantly larger than that of the neutral instructions conditioning group. It is the purpose of the present investigation to replicate the facilitory and inhibitory conditions of Hill with the additions of a pseudoconditioning group and a neutral instructions group so that the two primary questions listed earlier may be answered.

---

[1] This research was partially supported by Grant MH16911 awarded to the senior author from the United States Public Health Service. The authors wish to thank Robert Steward and Harry Coffey for their technical assistance.

[2] Requests for reprints should be sent to Michael E. Dawson, Department of Psychology, California State College at Los Angeles, 5151 State College Drive, Los Angeles, California 90032.

[3] The authors are also at the Research Department, Gateways Hospital, Los Angeles, California 90026.

METHOD

Subjects.—Eighty-six students enrolled in undergraduate psychology classes served as *S*s. The data of 6 *S*s are reported separately on the basis of a postexperimental inquiry. The remaining 80 *S*s were assigned to one of four groups by the method of block randomization. However, due to a procedural error which was not discovered until after data collection was completed, the number of *S*s per group was not kept equal. There were 20 *S*s in the facilitory group, 21 in the inhibitory group, 19 in the neutral group, and 20 in the pseudoconditioning group.

Apparatus.—The CS, a 1,000-Hz. tone of 50 db., was produced by a Maico audiometer and delivered through earphones. The UCS, a dc shock, was administered through 1.7-cm. circular silver electrodes attached to the tip of the little finger and the back of *S*'s left hand. The CS duration of 600 msec., the UCS duration of 100 msec., and the CS-UCS interval of 500 msec. were controlled by a Stoelting three-channel timer. The intertrial intervals of 30, 45, and 60 sec., with an average of 45 sec., were manually controlled according to a preestablished block randomized order.

The GSR was obtained as a dc resistance change through 1.3 × 1.6 cm. silver electrodes and recorded on a Grass Model 5 polygraph. The electrodes were coated with Burdick electrode paste and attached to the first and third fingertip of *S*'s right hand. Current through *S*'s fingers was 50 μa.

Design and procedure.—Four independent groups were employed. Three of the groups were administered CS-UCS pairings, while the fourth group was presented pseudoconditioning trials. The three conditioning groups differed in terms of the instructions presented: facilitory, inhibitory, or neutral. The experiment consisted of five parts:

1. Instructions. As in Hill's (1967) experiment, the instructions were given in two phases. First, the three conditioning groups were presented a rather detailed review of classical conditioning and a description of the CS and UCS used in the present experiment. Second, the facilitory and inhibitory groups were told that they could control their responses through their attitudes. The facilitory group was then told that it was "the most adaptive, sensible, and intelligent thing *to become conditioned*," while the inhibitory group was told that it was "the most adaptive, sensible, and intelligent thing *to not become conditioned.*" The neutral instructions group was not given the second phase of the instructions. The pseudoconditioning group was merely told that tone and shock would be presented randomly.

2. Shock alone. All *S*s were administered three shock-alone trials at the intensities of 1.0, 1.2, and 1.4 ma.

3. Acquisition. The three conditioning groups were presented 24 acquisition trials with 67% reinforcement. Nonreinforced trials occurred on Trials 3, 5, 9, 10, 13, 15, 16, and 21. The intensity of the UCS was gradually increased until it reached its maximum of 2.5 ma. on Trial 14. During acquisition the pseudoconditioning group was presented unpaired CSs and UCSs in such a manner as to equate (*a*) the number of CSs, (*b*) the number, intensity, and interval between UCSs, and (*c*) the total testing time with that of the three conditioning groups.

4. Extinction. All *S*s were administered five presentations of the CS without the UCS.

5. Postexperimental inquiry. All *S*s in the facilitory group were asked whether they tried to become conditioned and all *S*s in the inhibitory group were asked whether they tried to avoid becoming conditioned.

CR definition.—The largest change in resistance with a latency of 1.0 to 3.5 sec. after the onset of the CS during a test trial was recorded as the CR. The resistance scores were transformed to the square root of the change in conductance and the logarithm of the change in conductance. Since the square root and the logarithm transformations yielded the same conclusions, only the square root data are reported.

RESULTS

Shock alone.—A Duncan's range test with correction for unequal *N*s (McGuigan, 1968) demonstrated that the UCR magnitude of the four experimental groups did not significantly differ on the first shock-alone trial. This finding supports the experimental presupposition that the four groups did not differ in their initial GSR responsivity.

Acquisition.—The mean GSR acquisition data are presented in Fig. 1. A weighted-means analysis of variance revealed a significant group effect, F (3, 76) = 14.51, $p < .001$, with no trial or Trial × Group interaction effects. A Duncan's range test with correction for unequal *N*s demonstrated: (*a*) the GSR magnitude of the three conditioning groups was significantly larger than that of the pseudoconditioning group ($p < .01$); (*b*) the GSR magnitude of the facilitory group was significantly larger than that of the inhibitory group ($p < .05$); and (c) the GSR magnitude of neither the facilitory group nor the inhibitory group differed significantly from that of the neutral instructions groups. In order to determine the percentage of individual conditioners within each group, the following procedures were employed: First, each *S*'s mean GSR magnitude during acquisition was computed; second, the difference between each individual's mean and the pseudoconditioning group's mean during acquisition was computed; and third, the difference between each individual's mean and the pseudoconditioning group's mean was compared with the "least significant range" (McGuigan, 1968, p. 208). If the difference between the individual's mean and the pseudoconditioning group's mean was larger than the "least significant range," then it was concluded that the individual in question did condition. Using this criterion, it was found that 100% (20 of 20) of the facilitory group, 84% (16 of 19) of the neutral group, and 76% (16 of 21) of the inhibitory group conditioned.

Extinction.—The mean GSR extinction data are presented in Fig. 2. A weighted-means analysis of variance revealed a significant group effect, $F\ (3, 76) = 15.71$, $p < .001$, and trial effect, $F\ (4, 304) = 5.33$, $p < .01$, with no interaction effect. A Duncan's range test with correction for unequal *N*s demonstrated the same group differences to be significant during extinction as were presented during acquisition. Using the criterion for inferring the percentage of individual conditioners described earlier, it was found that 100% (20 of 20) of the facilitory group, 68% (13 of 19) of the neutral group, and 71% (15 of 21) of the inhibitory group retained the CR during the extinction trials.

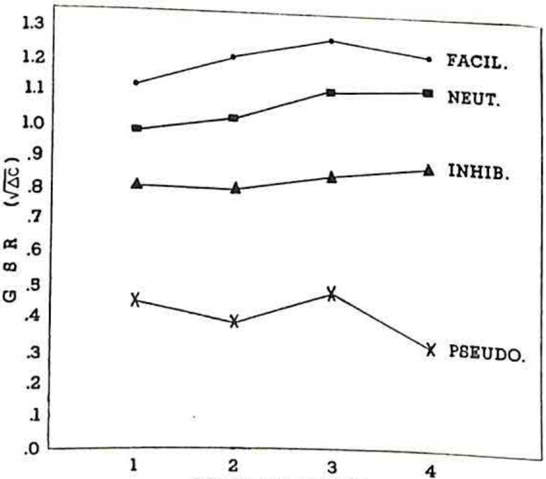


FIG. 1. Mean GSR magnitude on pairs of acquisition test trials.

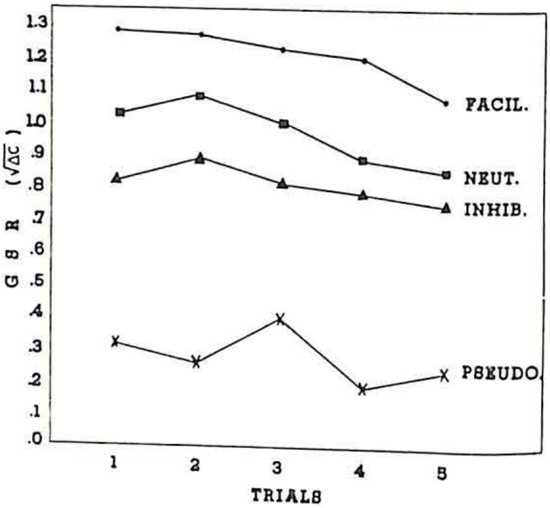


FIG. 2. Mean GSR magnitude on extinction trials.

Postexperimental inquiry.—During the postexperimental inquiry, two *S*s in the facilitory group stated that they did not try to become conditioned and four *S*s in the inhibitory group stated that they did not try to avoid becoming conditioned. Although this sample size is too small to permit reliable statistical inference, it should be noted that the mean GSR magnitudes during acquisition and extinction of all six "non-triers" were larger than the means of the neutral instructions group during acquisition and extinction.

UCR habituation.—In order to test the hypothesis that the difference in conditioning between the groups is associated with differential UCR habituation, the difference between the first UCR magnitude and the last UCR magnitude was computed for all *S*s. A Duncan's range test with correction for unequal *N*s was applied to these difference scores. It was demonstrated that (a)

UCR habituation for the inhibitory group was significantly greater than that for both the facilitory group ($p < .05$) and the neutral group ($p < .05$), and (*b*) UCR habituation for the facilitory group was less than that of the neutral group, although not significantly so.

DISCUSSION

The primary purpose of the present investigation was to determine whether (*a*) inhibitory instructions can completely inhibit GSR conditioning and (*b*) facilitory instructions can significantly facilitate GSR conditioning. The presently reported data indicate that (*a*) inhibitory instructions do not completely inhibit GSR conditioning, although they do tend to decrease the level of conditioning when compared to neutral instructions, and *b*) facilitory instructions do not significantly facilitate GSR conditioning, although they do tend to increase the level of conditioning when compared to neutral instructions. Taken together, the present results confirm the basic finding of Hill (1967), but suggest that her report of a significant set effect on GSR conditioning was due to the summation of facilitory and inhibitory effects rather than to the individual effects of each set. The present results are also compatible with reports which indicate that in nonconditioning situations, autonomic responses can be modified by instructional sets (Stern & Lewis, 1968; Sternbach, 1964).

The present results dealing with the effects of set on GSR conditioning share several similarities with the reported effects of set on eyelid conditioning. One similarity is that inhibitory set has been found to not completely abolish conditioning (Norris & Grant, 1948). A second similarity is that set has been found to produce an immediate effect on CR strength (Kimble, 1967, p. 645). A third similarity is that the difference in conditioning between facilitory and inhibitory groups is associated with differential UCR habituation (Nicholls & Kimble, 1964). A dissimilarity is that inhibitory set and facilitory set have been found significantly to inhibit and to facilitate conditioning, respectively (Hilgard & Humphreys, 1938; Norris & Grant, 1948). However, before this dissimilarity between GSR and eyelid conditioning is interpreted as a basic difference in the response systems, it should be emphasized that instructions used to establish sets in eyelid conditioning experiments have been stated in very explicit terms ("blink" and "do not blink"), while the instructions used in GSR conditioning experiments are stated in implicit terms: "It is adaptive, sensible, and intelligent to become conditioned," and "It is adaptive, sensible, and intelligent to not become conditioned."

A frequent theoretical interpretation of the effect of instructional sets on classical conditioning is as a source of drive level (Hill, 1967; Lindley & Moyer, 1961; Wickens, Allen, & Hill, 1963). However, before any adequate explanation is developed, the possibility of skeletal mediation must be systematically investigated. Katkin and Murray (1968) discussed this problem in regard to the distinction between "controlling" and "conditioning" autonomic responses, and much of what they had to say is relevant to the question of set effects. Although all *S*s in the present study were told to sit still and none of the *S*s reported using skeletal mediation, it is still possible that small changes in muscle tension and respiration may have contributed to the differences in GSR between the facilitory and inhibitory groups.

REFERENCES

BRIDGER, W. H., & MANDEL, I. J. A comparison of GSR fear responses produced by threat and electric shock. *Journal of Psychiatric Research*, 1964, **2**, 31–40.

COOK, S. W., & HARRIS, R. E. The verbal conditioning of the galvanic skin response. *Journal of Experimental Psychology*, 1937, **21**, 202–210.

DAWSON, M. E., & GRINGS, W. W. Comparison of classical conditioning and relational learning. *Journal of Experimental Psychology*, 1968, **76**, 227–231.

GRINGS, W. W., & LOCKHART, R. A. Effects of "anxiety-lessening" instructions and differential set development on the extinction of the GSR. *Journal of Experimental Psychology*, 1963, **66**, 292–299.

HILGARD, E. R., & HUMPHREYS, L. G. The effect of supporting and antagonistic voluntary instructions on conditioned discrimination. *Journal of Experimental Psychology*, 1938, **22**, 291–304.

HILL, F. A. Effects of instructions and subject's need for approval on the conditioned galvanic skin response. *Journal of Experimental Psychology*, 1967, **73**, 461–467.

KATKIN, E. S., & MURRAY, E. N. Instrumental conditioning of automatically mediated behavior: Theoretical and methodological issues. *Psychological Bulletin*, 1968, **70**, 52–68.

KIMBLE, G. A. (Ed.) *Foundations of conditioning and learning.* New York: Appleton-Century-Crofts, 1967.

LINDLEY, R. H., & MOYER, K. E. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *Journal of Experimental Psychology,* 1961, **61**, 82–88.

McGUIGAN, F. J. *Experimental psychology.* (2nd ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1968.

NICHOLLS, M. F., & KIMBLE, G. A. Effect of instructions upon eyelid conditioning. *Journal of Experimental Psychology,* 1964, **67**, 400-402.

NORRIS, E. B., & GRANT, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *American Journal of Psychology,* 1948, **61**, 37–49.

STERN, R. M., & LEWIS, N. L. Ability of actors to control their GSRs and express emotions. *Psychophysiology,* 1968, **4**, 294–299.

STERNBACH, R. A. The effects of instructional sets on autonomic responsivity. *Psychophysiology,* 1964, **1**, 67–72.

WICKENS, D. D., ALLEN, C. K., & HILL, F. A. Effect of instructions and UCS strength on extinction of the conditioned GSR. *Journal of Experimental Psychology,* 1963, **66**, 235-240.

(Received April 1, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 467–471

ABSTRACTNESS AND RECALL OF CONNECTED DISCOURSE[1]

JOHN C. YUILLE[2] AND ALLAN PAIVIO
University of British Columbia *University of Western Ontario*

The effects of content abstractness on the recall of connected discourse were examined. Groups of *S*s received two learning and recall trials with a list of 79 words at one of three levels of content abstractness. Half of the *S*s were given the words in syntactic order; the remainder saw the words in a random order. Abstractness affected total recall scores and the type of word recalled. Level of organization (syntactic vs. random) affected recall and the number and type of errors, especially on Trial 2. A triple interaction indicated that recall was facilitated by syntactic order on the second trial only when the content of the passage was highly concrete. The results were interpreted as consistent with the hypothesis that thematic storage is in the form of imagery.

A number of recent studies have investigated factors affecting the retention of connected discourse. Lachman and his associates (Lachman & Dooling, 1968; Pompi & Lachman, 1967) have hypothesized that the theme in connected discourse generates surrogate processes reflecting the essential idea of a passage and that use of these processes aids in regeneration of the passage. Pompi and Lachman supported this hypothesis by finding more thematically related errors in recognition and recall of connected discourse than of randomly presented material. In considering the processes by which *S* might store the discourse theme, they suggested visual images. Thus, "the probability of a surrogate structure being generated" would be reduced "if the original theme was sufficiently abstract [p. 144]." The present study was an attempt to investigate this possibility.

A series of studies by Paivio and his associates (for a summary, see Paivio, 1969) has concentrated on the role of visual imagery in verbal learning and memory. The rated capacity of nouns to evoke imagery (I) has been found to be positively related to paired-associate learning (e.g., Paivio, Yuille, & Smythe, 1966) and to serial and free recall (Paivio, Yuille, & Rogers, 1969). These results have supported the hypothesis that images can function as effective mediators of response recall. Imagery may serve a similar role in the recall of connected discourse, as Pompi and Lachman (1967) suggest. The reader may store the "basic idea" of a paragraph in the form of a visual image. When asked to recall the verbal content of the paragraph, *S* generates verbal associations to the stored image. In order to determine the possible role of imagery in the recall of connected discourse, the present study involved obtaining recall for paragraphs of varying degrees of abstract content.[3] The words in the paragraph were organized for presentation in two ways, syntactic and random.

Two predictions concerning the results were made. First, from the earlier work of Paivio and his associates (see Paivio, 1969), a main effect of abstractness was expected, with higher recall for the words of concrete than abstract content paragraphs. This main effect should be reflected particularly in the recall of nouns since nouns tend to reflect abstractness more than other word types. Conversely, recall of function words (e.g., "of," "from," etc.) should be unrelated to abstractness. An Organization × Abstractness interaction was also predicted. If imagery is used as a memory aid for thematic material, then higher recall should be obtained with concrete material in the syntactic than in the random condition, while organization should have less of an effect with abstract content.

---

[1] This study was completed while the first author held a National Research Council of Canada postdoctoral fellowship at McGill University. The authors wish to thank Rhona Katz for assistance in the preparation of materials and for conducting a related pilot investigation.

[2] Requests for reprints should be sent to John C. Yuille, Department of Psychology, University of British Columbia, Vancouver 8, Canada.

[3] The abstractness of the content was varied in the present study, which may be considered the same as varying theme abstractness. However, combinations of abstract units might suggest a concrete theme. Consequently, although *S*s rated the themes in terms of abstractness, the defining variable, operationally, is abstractness of content.

METHOD

Subjects.—Ninety-six male, Grade 10 high school students served as *S*s.

Materials.—Six 79-word paragraphs were used, two of concrete content, two with moderately abstract content, and two with abstract content. The "concrete" paragraphs (designated C_1 and C_2) were taken from Pompi and Lachman (1967). Both paragraphs are descriptions of concrete scenes; C_1 describes a medical operation, C_2 a battle scene. The "moderate abstract" paragraphs (M_1 and M_2) and the "abstract" paragraphs (A_1 and A_2) were constructed to be thematically similar to C_1 and C_2 but more abstract and general.

> C_1: Chief Resident Jones adjusted his face mask while anxiously surveying a pale figure secured to the long gleaming table before him one swift stroke of his small sharp instrument and a thin red line appeared then an eager young assistant carefully extended the opening as another aide pushed aside glistening surface fat so that vital parts were laid bare everyone present stared in horror at ugly growth too large for removal he now knew it was pointless to continue.

> C_2: Major Smith watched horrified as withering fire crumpled an immediate superior now he must guide the small group on their important journey urging his men to keep silent they moved through dense snake infested forests suddenly clattering caterpillar treads sounded nearby after rapidly cranking in one round he estimated distance and elevation his finger carefully tightened and the steel monster was blown apart fast thinking saved them although this had been only a minor incident among endless desperate encounters.

> M_1: Details concerning a torturous disease are seldom discussed by the general public medical personnel however cannot shelter themselves from such unpleasant malignancies at all times they must bear witness to internal growths which necessitate immediate attention chief residents inspect affected areas carefully using very sharp instruments to reach below body surface assistants and aides gathered round with intention of expanding scientific knowledge survey what has been uncovered often these findings reveal the futility and hopelessness of attempts in reducing agony.

> M_2: As a consequence of armed conflict leaders with subordinates undergo many serious hardships often plagued by extreme hunger or fatigue men must nonetheless march to face severe strife some may fall along the way while others meet this fate during actual struggle on the field friends are subject to continually watching comrades perish and endure much pain still instruments of destruction mutilate their helpless bodies and weakened minds all scars mark minor incidents among endless desperate but heroic encounters.

> A_1: Greatest problem confronting medical research is incurable growths that cannot be cut out because such agony results from these tissues ethical and moral issues need no consideration philosophical questions must never act as barriers against relieving pain discovering solutions to remove evils of disease remains paramount for our needs armchair talkers should stop heedless interference which only allows torture to continue truth lies in action preventing destruction in nature rather than debate concerning minor ideas of rights and wrongs.

> A_2: Despite honest intentions of many countries to avoid conflicts disagreements occur when different national interests clash contrary effects arise resulting events involve confrontations where senseless happenings take place while men sacrifice their lives for certain ideals they must temporarily suspend other basic principles regarding humanity and justice as a consequence of this intergroup strife normal life aims become subordinated to active struggle towards conquest for power individuals in serving military goals meet painful encounters effecting small gains and successes.

The four new paragraphs were written to conform as much as possible to the criteria used by Pompi and Lachman (1967). Each contained 75 different words plus 4 syntax words that occurred twice. A similar distribution of frequencies

TABLE 1

MEAN RATINGS AND FREQUENCY DISTRIBUTIONS OF WORDS IN EACH OF SIX PARAGRAPHS

Paragraph	Emotionality	Concreteness	Specificity	Frequency	
				A-AA	10 or less
C_1	5.33	6.83	7.00	56	4
C_2	4.83	6.83	6.50	58	6
M_1	4.83	5.17	3.50	54	7
M_2	5.33	4.83	4.50	49	6
A_1	5.33	2.83	4.33	55	6
A_2	4.33	2.50	1.83	49	10

of occurrence (Thorndike & Lorge, 1944) of the words in the paragraphs was maintained (see Table 1). In addition, six judges rated each of the six paragraphs on three, seven-point semantic-type scales. The end points of the scales were defined as follows: (*a*) 7-specific, 1-general; (*b*) 7-high emotion, 1-low emotion; (*c*) 7-concrete, 1-abstract. The average ratings are presented in Table 1.[4]

Following the procedure of Pompi and Lachman (1967), each word of a paragraph was written on a blank IBM card. Four copies of each paragraph were made, making 24 separate decks of cards. Two of the decks of each paragraph were shuffled to be used for the random presentation condition.

Procedure.—The *S*s were tested in groups of 24 during regular classroom periods. Each *S* received one of the 24 decks of 80 IBM cards; the top card contained a label for his experimental condition. The *S*s also received a set of mimeographed instructions and a two-page response sheet. Instructions were the same for all *S*s, except that those receiving the randomly ordered material were instructed that the words were randomly drawn from the dictionary. The Ss receiving the material in syntactic order were informed that the words they would learn formed a short story. All *S*s were instructed to look at each word and try to learn it. They proceeded from one word to the next to the pace of an electric metronome at a rate of 2 sec/word. A 1-min. rest period was given after the last word was viewed, and then *S*s were given 4 min. to recall the words in any order. The presentation and recall procedure was then repeated.

During each testing session, four *S*s received one of the six paragraphs; two *S*s saw the paragraph in the syntactic order and two in the random order. Thus each group represented two replications of the experimental design.

RESULTS

The recall differences between the two different themes (medical vs. battle) were not considered in the analysis. The results were analyzed in a $3 \times 2 \times 2$ design with three levels of abstractness (concrete, medium, and abstract), two levels of organization (random vs. syntactic), and two trials.

The mean total numbers of words recalled in each condition are presented in Table 2. An analysis of variance revealed significant main effects of trials, $F\ (1, 90) = 214.55$, $p < .01$, and abstractness, $F\ (2, 90) = 6.65$, $p < .01$. The latter effect and a two-way Abstractness × Presentation Method interaction, $F\ (2, 90) = 3.36$, $p < .05$, were overridden by the significant three-way interaction, $F\ (2, 90) = 6.00$, $p < .01$. A comparison of the cell means by *t* tests revealed that the recall for Trial 1 did not vary as a function of experimental conditions. On Trial 2, significant differences were obtained only with syntactic presentation. The recall of the concrete content reliably exceeded recall of the medium and abstract content words, $t\ (30) \geq 3.05$, $p < .01$. Abstractness did not affect recall of the randomly presented material.

In order to determine the basis of the recall differences, recall of nouns and of function words (i.e., "of," "the," "at," "from") was examined separately. An analysis of variance was performed on the proportion of the total number of nouns in each paragraph recalled by each *S*. The mean values appear in Table 2. Significant main effects of abstractness, $F\ (2, 90) = 7.35$, $p < .01$, and trials, $F\ (1, 90) = 85.19$, $p < .01$, were overridden by an interaction of these variables, $F\ (2, 90) = 4.21$, $p < .05$. Regardless of the method of presentation, a greater proportion of nouns was recalled for the concrete and medium lists than for the abstract on Trial 1, while recall for the concrete list was superior to the medium and abstract groups on Trial 2. An inspection of the means reveals that these effects were largely due to the disproportionate increase in recall of the concrete list nouns on Trial 2.

---

[4] Meaningfulness (*m*) of the words in each paragraph was not equated. Sufficient evidence exists to conclude that *m* does not affect the learning of familiar words. For example, Paivio, Smythe, and Yuille (1968) conclude that "noun *m* has no positive effect whatever on PA learning when [imagery] is controlled [p. 438]." Similarly, Paivio et al. (1969) found only weak effects of noun *m*, relative to imagery, in free recall. The conclusion that *m* is a relatively ineffective variable is also supported by Kanungo (1968), Paivio (1968), and Paivio et al. (1966). Nevertheless, in order to check the relative *m* values of the component words of each paragraph, post hoc ratings of *m'* were obtained for each word. Twelve *S*s rated each word on a seven-point (1-low meaningful, 7-high meaningful) scale. The instructions were a modified version of those used by Noble, Stockwell, and Pryer (1957). The mean *m'* ratings for each paragraph (for 12 *S*s and 79 words) are: C_1, 2.96; C_2, 2.89; M_1, 3.08; M_2, 3.15; A_1, 3.06; A_2, 3.38. If anything, the rated *m'* of the concrete paragraph is somewhat lower than that of the more abstract ones.

TABLE 2

Recall and Errors as a Function of Content Abstractness, Organization, and Trials

Item	Abstractness					
	Concrete		Medium		Abstract	
	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
Mean total recall						
Syntactic order	12.25	22.83	11.38	17.19	9.63	14.38
Random order	11.75	17.38	11.50	19.25	10.06	16.63
Proportion of nouns recalled						
Syntactic order	.19	.38	.19	.24	.10	.18
Random order	.12	.23	.13	.22	.09	.21
Proportion of function words recalled						
Syntactic order	.25	.31	.22	.30	.16	.23
Random order	.31	.40	.35	.46	.32	.31
Proportion of theme-related errors						
Syntactic order	.33	.25	.39	.31	.55	.37
Random order	.16	.23	.14	.20	.18	.26

A main effect of order, $F\ (1, 90) = 6.96$, $p < .01$, revealed that a greater proportion of nouns was recalled in the syntactic order.

The proportion of total function words recalled was also determined (see Table 2). An analysis revealed two significant main effects: order, $F\ (1, 90) = 23.01$, $p < .01$, indicating that a greater proportion of function words was recalled in the random order, and a trials effect, $F\ (1, 90) = 21.94$, $p < .01$, with a greater proportion being recalled on Trial 2.

The final analysis concerned the number of intrusions made by *S* during recall. Analysis of variance of the total number of intrusion errors indicated that more errors were made on Trial 2, $F\ (1, 90) = 9.74$, $p < .01$, and less errors were made in the random order, $F\ (1, 90) = 6.97$, $p < .01$. The effect of theme is revealed in the proportion of the errors that are related to the theme. Two judges classified each error as theme-related or not theme-related. The judges agreed on the classifications of 86% of the errors. The proportion of the errors related to theme are given in Table 2. An analysis of these data revealed that *S*s given the words in thematic order made more theme-related errors, $F\ (1, 90) = 11.15$, $p < .01$. An Order × Trials interaction, $F\ (1, 90) = 5.69$, $p < .05$, however, indicates that the order effect occurred only on the first trial, and the thematic and random groups have the same proportion of errors on Trial 2.

Discussion

First-trial recall was not affected by the experimental conditions. This differs from the Pompi and Lachman (1967) finding that organization affected recall after a single presentation. In the present study, however, second-trial recall of the concrete material was higher with syntactic than random organization, a comparable result to Pompi and Lachman's. The difference in results between the two studies thus appears to be minimal.

The analysis of theme-related errors also revealed a pattern similar to that obtained by Pompi and Lachman (1967). On the first trial, more theme-related errors were made by the syntactic than the random order group. Pompi and Lachman suggested that this type of pattern reflects *S*s' guessing strategies during recall. The *S*s presented the list in syntactic order make guesses on the basis of the theme and consequently make more theme-related errors. The failure to find a similar effect on Trial 2 of the present study does not conflict with this interpretation. The second exposure to the list gives *S* more familiarity with the list words, decreasing the likelihood of making theme-related errors.

Consistent with other verbal learning studies, the main effect of abstractness, obtained in the analyses of total recall and the proportion of nouns recalled, adds further evidence to the potent effect abstractness exerts on learning. Although these results are not unexpected, they extend the previous findings to a new type of learning situation. In addition, the fact that abstractness had such a large effect on the recall results in the present study argues for imagery as the mediator of recall.

The major question of the present investigation was the extent to which imagery functions as the means of theme storage. Recall on Trial 2 was highest for the concrete syntactic condition and lowest for the abstract syntactic group. In fact, the second-trial differences due to method of presentation are solely a function of the high recall in the concrete syntactic condition. Thematicity had no effect on the recall of abstract material. This effect of abstractness on recall was also reflected in the proportion of nouns recalled. Abstractness, except for trials, accounts for most of the variance in performance.

It is possible that other word attributes may have influenced the results, although this seems unlikely. Meaningfulness does not appear to be systematically related (see Footnote 4). A recent study by Craik (1968) appears to rule out word length as an effective variable for either primary memory or secondary memory components of free recall. The findings of the present study, then, suggest that imagery may play an important role in the storage of a theme. Perhaps imaginal representation of the theme is the basis of the surrogate processes suggested by Pompi and Lachman (1967). Because of the limitations of the present study, however, a considerable amount of further research is required before any clear statement can be made about the nature of mediating processes in the learning of connected discourse.

REFERENCES

Craik, F. I. M. Two components in free recall. *Journal of Verbal Learning and Verbal Behavior,* 1968, 7, 996–1004.

Kanungo, R. Paired-associate learning of function words. *Psychonomic Science,* 1968, **10**, 47–48.

Lachman, R., & Dooling, D. J. Connected discourse and random strings: Effects of number of inputs on recognition and recall. *Journal of Experimental Psychology,* 1968, 77, 517–522.

Noble, C. E., Stockwell, F. E., & Pryer, M. W. Meaningfulness (m') and association value (a) in paired-associate syllable learning. *Psychological Reports,* 1957, **3**, 441–452.

Paivio, A. A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning and Verbal Behavior,* 1968, 7, 41–49.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review,* 1969, **76**, 241–263.

Paivio, A., Smythe, P. C., & Yuille, J. C. Imagery versus meaningfulness of nouns in paired-associate learning. *Canadian Journal of Psychology,* 1968, 22, 427–441.

Paivio, A., Yuille, J. C., & Rogers, T. B. Noun imagery and meaningfulness in free and serial recall. *Journal of Experimental Psychology,* 1969, 79, 509–514.

Paivio, A., Yuille, J. C., & Smythe, P. C. Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology,* 1966, 20, 362–377.

Pompi, K. F., & Lachman, R. Surrogate processes in the short-term retention of connected discourse. *Journal of Experimental Psychology,* 1967, 75, 143–150.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received April 1, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 472–481

DIFFERENTIAL EFFECTS OF INCIDENTAL TASKS ON THE ORGANIZATION OF RECALL OF A LIST OF HIGHLY ASSOCIATED WORDS[1]

THOMAS S. HYDE AND JAMES J. JENKINS[2]

University of Minnesota

In three experiments involving 17 groups, the amount and organization of recall of word lists varied with the type of incidental task performed by *S*s during presentation of the list. All *S*s heard a randomized list of high-strength primary word associates. When the incidental task required using the word as a semantic unit (rating the word as to its pleasantness), recall and organization were equivalent to those of a control group with no incidental task. When the incidental task involved the word as an object (checking for certain letters or estimating the number of letters in the word), recall and organization were greatly reduced. The effects were unaltered by incidental-plus-recall instructions, doubling presentation time, and presenting the list twice.

The value of the free recall situation for studying the organization of verbal materials in memory is well known. Substantial differences in amount of recall can be produced by varying stimulus materials and, ordinarily, these differences can be attributed to differences in the organization imposed by *S*s on the material recalled. Such recall-facilitating organization is variously referred to as secondary organization, subjective organization, clustering, etc. Reviews of this literature may be found in Cofer (1965), Mandler (1967), and Tulving (1968).

In general, it appears that meaningful relations (taxonomic classes, logical classes, and prior associative relations) lead readily to well-organized recall. Repetition of the materials in new random orders and successive recall attempts also lead to increasing degrees of organization even in seemingly unrelated materials.

A problem of particular interest now is the elucidation of the mechanisms which bring this organization about. If, along with Melton (1963), we conceptualize the memory sequence as trace formation, trace storage, and trace retrieval, a first step in choosing among mechanisms might be: Where in this input, storage, and output model shall we suppose the organization occurs?

Some studies have assumed that the organization takes place via mediation during the activity of recall itself. This view, which was explicitly advanced by Jenkins and Russell (1952), is compatible with Deese's (1959) work and is echoed in Cofer's (1965) review. It holds that once a word is recalled, it acts as an auxiliary stimulus to help elicit another word that was on the list.

Tulving (1962, 1964, 1966), on the other hand, views such organization as a storage phenomenon, with the list of items being stored under higher order units corresponding to the relations existing in the stimulus lists. Organization in free recall reflects the organization within the memory storage itself.

Cohen (1966) and Slamecka (1968) suggested mechanisms of still a different sort. Slamecka, e.g., postulated that traces of individual items are stored independently, but that *S* also encodes some general representation of the list structure as a retrieval plan or strategy which guides the recall of the items. Highly organized stimulus lists produce greater recall and more organization because they are more amenable to some highly organized and efficient retrieval plan or strategy.

---

[1] This research was supported in part by the Center for Research in Human Learning. The Center is supported by the National Science Foundation (GS 1761), the National Institute of Child Health and Human Development (HD-01136), and the Graduate School of the University of Minnesota.

[2] Requests for reprints should be sent to James J. Jenkins, Center for Research in Human Learning, 400 Ford Hall, University of Minnesota, Minneapolis, Minnesota 55455.

Both Cofer (1965) and Tulving (1968) suggested that it might be profitable to vary the intraexperimental situation and look for differences in organization of recall while holding the stimulus materials constant. If large and consistent differences in organization can be produced with intraexperimental manipulations, it may be possible to separate the roles played by encoding, storage, and retrieval in the production of organization. The present study is an effort in that direction.

The experiments reported here were modeled on a study by Postman, Adams, and Bohm (1956) which compared the recall of high-strength primary associates in groups that performed incidental tasks during presentation of the words. The *S*s whose task was to rate the words as to frequency of usage in English showed lower recall than *S*s given standard recall instructions. Reduced recall was not, however, accompanied by any reduction in the relative amount of secondary organization (clustering of pairs).

Pilot work for the present study was similar in design to the Postman et al. (1956) experiment, but the results were quite different. Several types of incidental tasks were found to effect a marked reduction in both the organization and amount of recall. Other tasks had little or no effect on either dependent variable as compared to a control group which performed no incidental task.

The guiding hypothesis, arising from these pilot studies, was that clustering in recall depends on the use to which *S*s put the words during the incidental tasks. Tasks that involved meaningful interpretation of the stimulus materials (words as semantic units) were expected to have little or no effect on clustering, while tasks which required nonmeaningful treatment of the stimuli (words as objects) were expected to effect a reduction in organization (clustering) and a corresponding decrement in recall.

METHOD

In all experiments there were three basic groups: The recall only group simply heard the words, with the instructions that they were to recall the words later. The incidental group was given some task to perform on the stimulus materials during their presentation. The incidental + recall group performed the same task as that performed by the appropriate incidental group, but *S*s were also told that they would have to recall the words later. The major variable of interest within each experiment was a comparison of the amount and organization of the recall of the word list by the groups of *S*s who performed different types of incidental tasks during presentation of the words.

All *S*s were students in sections of a course in introductory laboratory psychology at the University of Minnesota. All experiments were conducted in the classroom. Each separate section made up one group. Assignment of students to sections was not systematic with respect to academic ability, college, major department, interest, aptitude, sex, or any other variable known to be relevant to the learning of verbal material.[3] Experimental conditions were randomly assigned to the sections, with the restriction that all seven incidental groups were run first, in order to keep the true nature of the experiment from becoming known to these *S*s.

Stimulus materials consisted of the 12 pairs of primary associates used by Jenkins, Mink, and Russell (1958) in their high-strength recall list. These were the Kent-Rosanoff pairs with the highest associative strength in the early Minnesota norms (Russell & Jenkins, 1954). The mean frequency of response for these pairs was 70.9%, with a range from 60.5% to 83.3%. The words were presented to *S*s in a random order, with the restriction that primary associates could not occur together in the list. All groups in the first two experiments were presented with a list of the same word order. In the third experiment, where the words were presented twice, the second reading of the list was a different randomization. All word lists contained four "filler" words, two at the beginning and two at the end. Fillers were selected from the Minnesota norms so that they showed no appreciable associative strength with any of the 24 stimulus words. Filler words were not counted in any data analyses.

All stimulus materials were recorded and presented to *S*s via tape recorder during the experiment. Within each experiment the same tape was used for all groups.

Although all experiments were conducted simultaneously and any condition may be compared to any other, for the sake of the reader, the findings are presented as three experiments, grouped on presentation conditions.

---

[3] Group assignment does not appear to have resulted in distorted findings. Pilot data on 350 *S*s who were randomly assigned to groups from the general *S* pool of introductory psychology students is in complete accord with the data presented here.

EXPERIMENT I

Seven groups of *S*s in the first experiment were presented with one reading of the stimulus words at a rate of one word every 2 sec. The *S*s in the recall-only group were read the list of words with the instruction that they would have to recall them as soon as *E* finished his presentation. The remaining six groups differed from each other in two respects. Three of them were incidental groups and three were incidental + recall groups. The former groups were not prewarned about the recall task, while the latter groups were. Within each of these conditions there were three different incidental tasks. It was hypothesized that one task (pleasant–unpleasant) would facilitate recall and clustering, while the other two tasks would adversely affect recall and clustering. The three tasks performed by *S*s were as follows:

1. The pleasant–unpleasant task. Two groups of *S*s rated the words as to their pleasantness or unpleasantness on a simple check list with one line for each stimulus word.
2. The "E" task. Two more groups were given the task of detecting Es in the words that were read to them. The *S*s had a check list with spaces for each stimulus word. They were directed to check if the word contained an E, otherwise to leave the space blank. Eight of the 24 stimulus words contained at least one E.
3. The number-of-letters task. Two other groups of *S*s estimated the number of letters in each word. They were told not to count letters, but to estimate the number and write it in the space provided for each word. The instructions for these same tasks to the subgroups in the incidental + recall condition were modified to indicate that *S* would have to recall the words when the task was finished. He was encouraged to do his best on both of the tasks and told that if he did not do his best on the first task, his results on the second (recall of the words) would be of little value. When the presentation of the list was finished, a set of standard recall instructions was read to all groups. The *S*s were then allowed 5 min. to recall as many words as they could. Very few *S*s recalled any words during the fourth and fifth minutes of the recall period.

Results and Discussion

Four different variables were analyzed:

1. Mean recall. This was the mean for each group of the total number of words recalled by each *S*, excluding the four filler words.

2. Percentage of clustering per opportunity. A cluster consisted of the recall of an associated pair together in either a forward or a backward order. Recall of one member of the pair constituted an "opportunity" for clustering to occur. If the other member of the pair immediately followed, the pair was scored as a cluster for that opportunity. If one member of the pair occurred alone, it was scored as an "opportunity," but not as a cluster. If the other member of the pair occurred later in recall, it was not scored as an opportunity for clustering. The score reported for each group was then the mean percentage of clustering per opportunity.

3. Mean categories. This and the following measure are the same as those used by Cohen (1966). A category was simply an opportunity for clustering. When a member of a pair occurred in recall, it was scored as one category, whether or not it was followed by its associate. If one member of a pair occurred alone in recall and was scored as a category, the later recall of the other member of the pair was not scored as a category. The value reported for each group was the mean of the number of such categories recalled by each *S* in the group.

4. Mean items per category (IPC). This was simply the mean items per category for *S*s in each group. This measure was the total recall for each *S* divided by the number of categories that he recalled. This measure differed from the clustering measure in that words did not have to occur together, i.e., clustered, to be counted as members of the same category.

Results for the first experiment are reported in Table 1. The means and standard deviations for each of the four variables are given for each group, as well as the number

TABLE 1

Means and Standard Deviations for All Variables for Exp. I

Variable	Group						
	Incidental			Incidental + Recall			Recall
	Pleasant–Unpleasant (1)	E (2)	Number of letters (3)	Pleasant–Unpleasant (4)	E (5)	Number of letters (6)	Control (7)
Mean recall	16.3	9.4	9.9	16.6	10.4	12.4	16.1
SD	3.1	3.5	3.5	3.1	4.0	3.9	3.9
Percentage of clustering	67.5	26.3	30.9	71.5	41.7	40.3	63.7
SD	23.7	23.8	22.8	21.8	27.1	27.8	28.4
Mean categories	8.9	6.3	6.5	9.0	6.2	7.6	9.1
SD	1.5	1.6	1.6	1.5	1.9	1.9	1.9
Mean IPC	1.8	1.5	1.5	1.9	1.6	1.6	1.8
SD	.2	.3	.3	.1	.4	.2	.2
N	43	32	39	29	38	39	37

of *S*s. In the table the groups are numbered from one to seven to facilitate the statistical analysis.

One-way analysis of variance indicated statistical significance for each of the four variables across the seven groups. All F values far exceed the .001 level of significance.

Mean recall.—Contrasts on individual cell means yielded consistent results across the four variables. Consider first the mean recall variable within the incidental condition. Group 1 (pleasant–unpleasant task) was significantly different from Groups 2 (Es) and 3 (number of letters), $F\ (1, 250) = 67.778$, $p < .01$; $F\ (1, 250) = 65.069$, $p < .01$. Groups 2 and 3 did not differ significantly from each other. Group 1 did not differ significantly from the recall-only group (Group 7). Both Groups 2 and 3 were significantly different from Group 7, $F\ (1, 250) = 59.718$, $p < .01$; $F\ (1, 250) = 56.151$, $p < .01$.

Results were almost identical for the mean recall within the incidental + recall condition. Group 4 (pleasant–unpleasant) differed significantly from Groups 5 (Es) and 6 (number of letters), $F\ (1, 250) = 50.241$, $p < .01$; $F\ (1, 250) = 23.064$, $p < .01$. Within this condition, Group 5 recalled fewer words than did Group 6, $F\ (1, 250) = 5.223$, $p < .05$. Group 4 did not differ from the recall-only group (Group 7). Groups 5 and 6 were both significantly different from Group 7, $F\ (1, 250) = 44.253$, $p < .01$; $F\ (1, 250) = 20.196$, $p < .01$.

Analyzing differences within tasks across the intentionality conditions, only the number-of-letters task showed a significant difference, $F\ (1, 250) = 9.176$, $p < .05$, indicating that recall prewarning had a positive effect. Differences between Groups 1 and 4 and between 2 and 5 were not significant.

Clustering.—Results for the second variable, percentage of clustering, were almost identical to those reported for the first variable, mean recall. Within the incidental subdivision the results were exactly the same as for the first variable. All significant F values far exceeded the .01 level of significance. Within the incidental + recall subdivision there was a return to the general regularity; Groups 5 and 6 were not significantly different as they had been with mean recall. In analyzing differences within tasks across the intentionality condition, results were also slightly different from the first variable. Groups 3 and 6 (number of letters) were not significantly different as they had been with mean recall. Instead, Groups 2 and 5 (Es) were significantly different, $F\ (1, 250) = 6.472$, $p < .05$. In all other respects the results were the same as for mean recall. All significant F values far exceeded the .01 level of significance.

Categories.—The results for the third variable, mean categories, followed the same pattern as the results from the first two variables. Within the incidental condition the results were identical to those found with the first two variables. The pleasant–unpleasant task yielded as many categories as the recall-only group, while the other two tasks resulted in fewer categories recalled. Within the incidental + recall condition Groups 5 and 6 were significantly different (as they had been on mean recall), F (1, 250) = 14.006, $p < .01$. No differences were found within tasks across the intentionality condition, i.e., Group 1 vs. 4 (pleasantness), 2 vs. 5 (Es), and 3 vs. 6 (number of letters) were not significantly different. In all other respects the results were the same as with the first two variables. All F values far exceeded the .01 level of significance.

Items per category.—The results for mean IPC followed the general pattern. Within the incidental condition, results were identical with those for the other three variables. The *S*s in the pleasant–unpleasant group recalled more items per category than did *S*s performing the other two tasks. Within the incidental + recall condition, results mirrored those found for the second variable (clustering), i.e., Groups 5 and 6 were not significantly different. As with the third variable, there were no differences within tasks across the intentionality conditions. In all other respects the results were exactly the same as for the other three variables. All F values far exceeded the .01 level of significance, except a comparison between Groups 4 and 5 which yielded an F (1, 250) = 6.750, $p < .05$.

Interpretation.—The results for the four variables showed a consistent pattern for the experiment. For all of the variables considered, the pleasant–unpleasant task was not appreciably different from the recall-only group, within either the incidental or the incidental + recall condition. The E task and the number-of-letters task were never different from each other in the incidental condition, but were different from each other within the incidental + recall condition for mean recall and mean categories. In all cases the E task and the number-of-letters task differed from both the recall-only group and the pleasant–unpleasant group. In most comparisons the tasks themselves did not have different effects on the groups in the incidental and the incidental + recall conditions. For mean recall, the number-of-letters task showed a small but significant difference, and for the percentage-of-clustering variable the E task showed a small but significant difference. For the last two variables, mean categories and mean IPC, none of the tasks showed a difference across these conditions.

Alternative analysis.—One possible explanation for the differences in organization and recall produced by the different tasks that immediately presented itself was that the two tasks, E and number of letters, that lowered organization were leading *S*s into some type of counterorganization during recall. To evaluate this hypothesis, the E condition was taken as a test. The recall characteristics of words containing the letter E were analyzed for the first experiment (see Table 2). Three measures were employed for this analysis: (*a*) recall of words containing an E; (*b*) mean recall of words containing an E as a percentage of total recall; and (*c*) percentage of clustering of words containing an E.

Results of these three variables are presented in Table 2. One-way analysis of variance on the first variable, recall of Es, yielded an F (6, 250) = 13.148, $p < .01$. Results indicated that the recall of Es mirrored exactly the mean total recall for each group. In all cases where the mean recall was significantly different between groups, so was the mean recall of words with an E. All differences exceeded the .01 level of significance.

When the recall of words containing an E was corrected for total recall, a much different pattern emerged. Groups 2 and 5 that performed the task of looking for Es now showed much higher recall of the words with E. One-way analysis of variance on the second variable, mean recall of words with E expressed as a percentage of total recall, yielded an F (6, 250) = 3.596, $p < .01$. Within the incidental condition, Group 2 (Es) was significantly greater than Groups 1, 3, and 7, F (1, 250) = 11.231, $p < .01$; F (1, 250) = 10.138, $p < .01$; F (1, 250) = 12.454, $p < .01$. Within the incidental +

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE ANALYSIS OF THE WORDS CONTAINING THE LETTER E FOR EXP. I

Variable	Group						
	Incidental			Incidental + Recall			Recall
	Pleasant–Unpleasant (1)	E (2)	Number of letters (3)	Pleasant–Unpleasant (4)	E (5)	Number of letters (6)	Control (7)
Mean recall	5.4	3.8	3.4	6.1	3.8	4.2	5.4
SD	1.7	1.4	1.8	1.4	1.7	1.8	1.9
Percentage of mean recall/total recall	32.7	42.2	33.5	36.1	39.1	33.5	32.9
SD	7.9	18.0	11.0	7.3	15.3	8.0	9.7
Percentage of clustering	50.9	40.4	41.3	53.6	45.5	42.8	44.8
SD	28.9	34.0	38.6	18.8	31.4	32.2	23.2
N	43	32	39	29	38	38	37

recall condition, Group 5 (Es) was significantly greater than Groups 6 and 7, F $(1, 250) = 4.623$, $p < .05$; F $(1, 250) = 5.025$, $p < .05$. Group 5 did not, however, differ significantly from Group 4.

Analysis of the third variable, percentage of clustering of words containing an E, yielded statistically unreliable results. The rather high percentage of clustering found in all groups (see Table 2) can be explained as being due to the fact that half of the words containing an E occurred in words that were primary associates.

Groups that looked for Es showed higher proportional recall of words with an E. In the organizational measure, however, there were no differences. The *S*s who performed the task of looking for Es did not show a significantly greater tendency to organize their recall around this dimension.

EXPERIMENT II

In the second experiment, *S*s were given a much longer time interval per word during stimulus input. The words were read at a rate of one word per 4 sec. rather than at one per 2 sec. as in the first experiment. This was the rate of presentation used by Postman et al. (1956) and was the major difference, other than the type of task, between our pilot work and Postman's experiment. This condition was run as both a systematic replication of the first experiment, and, also, as an attempt to clear up differences in results found between the pilot work and Postman's experiment.

The number-of-letters task was dropped from the study at this point. This made a total of five groups in all. The control group was similar in all respects to the control group in the first experiment, except that the words were read at a slower rate. The four remaining groups were made up of the two remaining incidental tasks, the pleasant–unpleasant task, and the task in which *S*s looked for letters. These were again subdivided into two subgroups across the intentionality condition. The task in which *S*s looked for letters was changed slightly to fill up the longer time interval; instead of checking the response sheet for the presence of the letter E, *S*s now looked for either Es or Gs and made a check if either or both of these letters were present in the word.

Instructions for all groups were changed slightly to accommodate the new conditions, and *S*s were told that they would have plenty of time for the task.

Results and Discussion

Results for the second experiment are reported in Table 3. The results are almost identical to those in the first experiment. One-way analysis of variance showed highly significant results with all four of the variables. All F values far exceeded the .001 level of significance.

TABLE 3

Means and Standard Deviations fot All Variables for Exp. II

Variable	Group				
	Incidental		Incidental + Recall		Recall
	Pleasant–Unpleasant (8)	E-G (9)	Pleasant-Unpleasant (10)	E-G (11)	Control (12)
Mean recall	17.0	8.9	18.2	11.5	16.6
SD	3.7	2.8	2.4	3.7	3.5
Percentage of clustering	78.8	17.3	68.7	38.0	67.4
SD	22.4	18.1	20.8	28.8	28.0
Mean categories	9.1	6.4	9.7	6.9	9.8
SD	1.2	1.8	1.1	1.7	1.9
Mean IPC	1.7	1.4	1.9	1.6	1.8
SD	.3	.3	.1	.3	.2
N	34	33	29	32	33

Comparisons of individual means yielded results that were consistent across the four variables and also consistent with the results for Exp. I.

Mean recall.—Within the incidental condition, Group 9 (E-G) was significantly different from both Groups 8 (pleasant–unpleasant) and 12 (recall only). Group 8 did not differ significantly from Group 12. Within the incidental + recall condition, the same pattern of differences emerged: Group 11 (E-G) was significantly different from Groups 10 (pleasant–unpleasant) and 12 (recall only), while Group 10 did not differ from Group 12. Within tasks across the intentionality conditions, the E-G task showed a significant difference, but Groups 8 and 10 (pleasant–unpleasant) were not significantly different. All significant *F* values were beyond the .01 level of significance.

Percentage of clustering.—Exactly the same pattern of results obtained for the second variable, percentage of clustering. All significant *F* values far exceeded the .01 level of significance.

Mean categories.—Results for the third variable, mean categories, were identical with the earlier findings except that the E-G task showed no difference across the intentionality conditions; i.e., the Group 9 vs. 11 comparison was not statistically significant. All significant *F* values far exceeded the .01 level of significance.

Mean IPC.—Results for the last variable, mean IPC, were the same in all respects as those for the first two variables. All significant *F* values far exceeded the .01 level of significance. It is again clear that the pattern is consistent for all four variables and the overall pattern is the same as for the first experiment.

Experiment III

In the third experiment, the conditions were the same as in the first experiment except that the word list was read twice to *S*s at a rate of one word per 2 sec. The first randomization of the list was the same as in the first two experiments; the second time through the list a second randomization was used. Treatments were arranged as in the second experiment. Instructions were the same as those in the first experiment except that *S*s were told that many of the words would be read more than once. Response sheets included spaces to respond to 52 words, rather than 28 as in the first two experiments.

The third experiment was primarily a systematic replication of the first two experiments, across a different intraexperimental variable (repetition) known to have an effect on clustering in recall.

Results and Discussion

Results for the third experiment, reported in Table 4, mirror those of the first two

TABLE 4

Means and Standard Deviations for All Variables for Exp. III

Variable	Group				
	Incidental		Incidental + Recall		Recall
	Pleasant–Unpleasant (13)	E (14)	Pleasant–Unpleasant (15)	E (16)	Control (17)
Mean recall	18.7	13.1	18.5	13.3	17.9
SD	2.2	2.7	3.0	4.8	2.5
Percentage of clustering	74.4	35.3	74.1	39.4	72.9
SD	17.9	23.1	19.0	25.3	24.9
Mean categories	10.0	7.9	10.0	8.0	9.6
SD	1.0	1.2	1.3	2.2	1.3
Mean IPC	1.9	1.7	1.9	1.6	1.9
SD	.1	.3	.2	.2	.2
N	30	33	34	27	40

experiments. All consistent findings of the earlier experiments were significant far beyond the .001 level.

Mean recall.—Results for the mean recall variable confirmed the earlier major findings. Group 14 (E) was significantly different from Groups 13 (pleasant–unpleasant) and 17 (recall only). Groups 13 and 17 did not differ significantly. Within the incidental + recall condition, the same pattern of results occurred; Group 16 (E) was significantly different from Groups 15 (pleasant–unpleasant) and 17 (recall only), while Group 15 did not differ significantly from Group 17. Differences within the tasks across the intentionality variable were not statistically significant.

Other variables.—The results for the other three variables were identical in all respects to those for mean recall. All significant *F* values far exceeded the .01 level of significance.

General Discussion

The major finding from these three experiments is that the tasks performed by *S*s while they listen to verbal stimuli greatly effect both amount of free recall and the apparent organization of the material recalled. Two of the tasks studied here, having *S*s look for the letter E in the words and having *S*s estimate the number of letters in the words, greatly reduced mean recall and the organization in recall as measured by percentage of clustering. The other task, rating words as pleasant or unpleasant, reduced neither recall nor organization in recall as compared to the recall-only group. Rather surprisingly, this result was obtained for both the incidental and the incidental + recall conditions. Furthermore, the effects attributable to the nature of the task were very large and reliable. For example, the differences in percentage of clustering between the pleasant–unpleasant and the E group within the incidental condition were 40% or greater.

The E and the number-of-letters task sometimes showed higher recall and higher percentage of clustering in the incidental + recall condition than they did in the incidental condition. These differences were small, however, and often failed to reach statistical significance.

One possible explanation for the differences in organization and recall produced by the different tasks is that the two tasks, E and number of letters, somehow led to some type of "counterorganization," which was inefficient. To check this hypothesis, the occurrences in recall of words containing the letter E were analyzed. It was found that *S*s showed no tendency to organize their recall around those words containing an E, so it appears that we must look elsewhere for an explanation of these differences.

What can be said about the differences in the tasks which produced such widely different results in organization in recall? Before a definitive division can be made between tasks that produce decrements in recall organization and tasks that do not, more research on many different types of tasks is needed. Current findings suggest to us that when *S* is performing some task where he is using the word as a semantic unit, organization in recall and, consequently, recall itself are not adversely affected. In the present experiments the pleasant–unpleasant task is of this type. The task used in the Postman et al. (1956) experiment where *S*s rated the words as to their frequency of usage in the English language appears also to be of this type.

What about the other "type" of task? In the other two tasks, the present *S*s appear to use words as objects or collections of letters rather than as units of meaning. This type of task seemed to lower the organization in recall consistently to a very great degree.

What do these results have to say with regard to the several mechanisms suggested to account for organization in recall? It seems that no matter how we characterize the differences between the tasks (the way *S* uses the word, amount of attention required of *S*, etc.) that produce differences in organization in recall, we are led to the same locus of the effect—the nature of the stored trace.

The first mechanism discussed was the notion that the organization in free recall is a primarily mediational phenomenon taking place during recall itself. The results of the present experiments do not lend support to such a postulation. During recall all the groups in the study were under essentially the same conditions. It is hard to explain, therefore, why all the groups did not show the same degree of organization in recall. This suggests, rather, that the groups were somehow unequal with respect to organization before the recall process began.

Cohen (1966) and Slamecka (1968) emphasize the input stage, suggesting that *S* is acquiring both the individual words and also some representation of the general structure of the list (Slamecka) or codes corresponding to the categories (Cohen). This seems more plausible since it deals with the locus of the present experimental manipulations, but still fails to be very persuasive as an explanatory device. The most damaging part of the present results to such an interpretation is the lack of any consistent difference in organization within the same tasks across the intentionality variable. The differences within the tasks across the incidental and incidental + recall conditions were small and more often than not failed to show statistical significance. The incidental + recall condition produced decrements in organization nearly equal to the incidental condition for the two tasks that interfered with organization. If the organization in recall is due to some encoding of a general representation of the list structure, the group that knew they were going to have to recall the words ought to have superior organization in recall. Even if such a process is covert and more or less automatic, any instructions that the words are to be recalled should produce a set in *S*s conducive to the operation of such a mechanism. (Incidentally, questioning of the 244 *S*s in the seven incidental groups after the experiment showed only 10 *S*s suspected that they might be required to recall the words.) The failure to find a difference associated with intention does not encourage us to believe we have a simple input variable.

The final suggestion as to a mechanism is that of Tulving (1962, 1966), who suggested that it is the traces of the words themselves that are stored together in some fashion that serves to facilitate organization in recall. The present work, like Tulving's, seems to fit with some type of storage suggestion more easily than with any other type of suggestion. The large differences between tasks, as well as the lack of differences along the intentionality condition, seems to suggest that differences in organization found in recall reflect differences in the way traces of individual items are arranged in storage. It is tentatively suggested that when words are used as units of meaning (the pleasant–unpleasant task, and the Postman et al., 1956, frequency estimation task), the semantic components of the words are activated. If the associates are strongly related semantically (and we know they are), this assures that common structures are activated in the task. Thus on the search for recall, the items to be recalled are found together. On the other hand, when words are used by *S*s as a collection of symbols devoid of meaning (at least as far as the task is concerned), the common structures are not activated and the recall is unorganized. The effect may be further augmented by some general superiority of semantic activity in processing materials for later recall.

REFERENCES

Cofer, C. N. Some factors in the organizational characteristics of free recall. *American Psychologist*, 1965, **20**, 261–272.

Cohen, B. H. Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 182–187.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, 1959, **5**, 305–312.

JENKINS, J. J., MINK, W. D., & RUSSELL, W. A. Associative clustering as a function of verbal association strength. *Psychological Reports,* 1958, **4**, 127–136.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *Journal of Abnormal and Social Psychology,* 1952, **47**, 818–821.

MANDLER, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

MELTON, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 1–21.

POSTMAN, L., ADAMS, P. A., & BOHM, A. M. Studies in incidental learning: V. Recall for order and associative clustering. *Journal of Experimental Psychology,* 1956, **51**, 334–342.

RUSSELL, W. A., & JENKINS, J. J. *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test.* (Tech. Rep. No. 11, Contract N8-ONR-66216) Washington, D. C.: United States Government Printing Office, 1954.

SLAMECKA, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology,* 1968, **76**, 504–513.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychological Review,* 1962, **69**, 344–354.

TULVING, E. Intratrial and intertrial retention: Notes towards a theory of free recall in verbal learning. *Psychological Review,* 1964, **71**, 219–237.

TULVING, E. Subjective organization and effects of repetition in multitrial free recall learning. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 193–197.

TULVING, E. Free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* New Jersey: Prentice-Hall, 1968.

(Received April 4, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 482–487

EFFECTS OF STIMULUS INFORMATION REDUCTION ON SEARCH TIME OF RETARDED ADOLESCENTS AND NORMAL CHILDREN

HERMAN H. SPITZ[1]

E. R. Johnstone Training and Research Center, Bordentown, New Jersey

Adolescent retardates and fourth- and seventh-grade normals were compared on a visual search task in which 16 target pieces ranged from high to low information. The target pieces were from a children's jigsaw puzzle. Eight were natural pieces judged for content information, and 8 were duplicate pieces painted flat gray and rated for contour information. As information in the target pieces declined, performance of the adolescent retardates and fourth graders rapidly deteriorated. The seventh graders coped more successfully with the reduced information.

It has been suggested that mental retardates have a reduced capacity to process incoming stimuli. In particular, it has been pointed out that adolescent retardates do not organize incoming information or develop strategies as effectively as do normals of equal chronological age (CA). Consequently, retardates are handicapped in their capacity to handle information (Spitz, 1966). Recent findings indicate that young children are similarly handicapped (Siegel, 1968) and that this problem is at least partially related to mental age (MA) (Spitz, 1967; Spitz, Hoats, & Holden, 1968).

Visual search involves the use of particular kinds of strategies. The *S* must actively abstract relevant features of the target stimulus in order to find its match in the embedding field (Neisser, 1967). Targets may be made relatively easy to find as, e.g., when a target containing a large amount of information is embedded in a field of stimuli containing substantially less information. However, as the information level of the target is reduced, depriving *S* of distinctive properties in the stimulus, search time increases. Conditions of reduced distinctiveness make increased demands on *S*'s capacity to seek out and use the remaining distinguishing features. Under these circumstances, retardates and young normal children, with fewer strategic resources for coping with reduced information input, should be more adversely affected by information deprivation than older normal children.

The present experiment tested this inference with a method suggested by Pollack and Spence (1968). Their approach, in turn, was stimulated by the work of Mackworth and his associates (cf. Mackworth & Morandi, 1967), who had found a positive relationship between the concentration of eye fixations and the rated amount of information in certain sections of a picture. Pollack and Spence reported that judged ratings of parts of a picture give a reliable measure of information value and can be used to predict performance on a visual search task.

METHOD

Material

The natural puzzle.—The puzzle (Fairy Tale Puzzles, Platt & Munk) was a 28-piece, 9 × 11 in. chromatic scene from Rumpelstiltskin. Pollack and Spence's (1968) instructions were used to obtain a measure of the judged information value of each puzzle piece. Five colleagues and two college students were requested to point out the piece contributing the most pictorial information to the entire picture and the piece contributing the least pictorial information. They then rated the remaining 26 pieces on a 12-point scale ranging from least to most informative. The average of the seven judgments provided a measure of the information value of each piece.

---

[1] The author is most grateful to Samuel Kaltovich, principal of Roebling School No. 5, Roebling, New Jersey, for invaluable assistance in providing space and students. Thanks are due also to the members of the Johnstone Research Department for critical readings of the manuscript.

Requests for reprints should be sent to Herman H. Spitz, Edward R. Johnstone Training and Research Center, Bordentown, New Jersey 08505.

The judgments were highly consistent, as evidenced by a Kendall coefficient of concordance (W) of 2.58 ($p < .001$) for rankings of the 8 pieces subsequently chosen as target pieces.

The blank puzzle.—To eliminate content information available in the pieces, a duplicate puzzle was painted flat gray, so that only shape was available to guide S in finding a target piece.

Since complexity has been described as a measure of information (Garner, 1962) and since number of turns is highly related to the judged complexity of meaningless shapes (Attneave, 1957), the number of turns of the contour of each puzzle piece was tabulated. A simplified adaptation of a method suggested by Attneave and Arnoult (1956) was used, in which a subjectively determined polygon was drawn around each piece. By counting every angle of 150° or less, a rough but serviceable measure of the complexity and, consequently, of the information value of the blank pieces was obtained.

The target pieces.—Four high-information natural puzzle pieces were selected, two of which had high-information contours and two of which had low-information contours. Four low-information natural pieces were also chosen, so that two had high- and two had low-information contours (see Table 2). These 16 pieces, 8 natural pieces and duplicates of these same 8 pieces painted flat gray, became the target pieces that Ss were required to match to corresponding pieces in the complete puzzle.

Apparatus and Procedure

The stimuli were presented in a modified Wisconsin General Test Apparatus. Two fluorescent desk lamps, one on each side of S's section of the apparatus, provided light and prevented S from seeing E through the one-way mirror. The complete puzzle, with a target piece centered directly above it, was slid in front of S for each trial.

Preliminary to the experiment, a sample puzzle (Cinderella) with an easy target piece in place was slid in front of S, and E pointed out the duplicate piece in the puzzle. As a second sample, the experimental blank puzzle was presented (upside down for the example only), and the correct match for a target piece (not used in the experiment proper) was pointed out by E. The S was told that these were samples of the types of puzzle to be used and was assured that the target piece would always be facing the correct way. He was instructed to find the target piece as quickly as he could and touch it lightly with the pointer, as E had done in the examples.

Timing by E on a .1-sec. stopwatch started when the sliding panel on which the puzzle rested reached the end of its excursion and terminated when S touched the correct piece. If S touched an incorrect piece, he was informed of his error, but timing was not stopped. As soon as the correct piece was touched, the puzzle was retracted and prepared for the next trial. If after 150 sec. of search, S had not touched the correct piece, E pointed it out and terminated the trial.

Half of the Ss of each group were given the natural puzzle first. Starting points were distributed throughout each group, and for each S the same order of presentation was followed for the blank as for the natural puzzle.

In sum, each S had 16 trials, 8 in which he searched for natural target pieces in the natural puzzle and 8 in which he searched for blank target pieces in the blank puzzle. It is unlikely that any Ss noticed that the eight blank pieces were identical to the eight natural pieces but with content obscured.

Subjects

Three groups of Ss were used: fourth- and seventh-grade students from a local grade school and adolescent institutionalized retardates in residence at the Johnstone Center. Characteristics of the groups are given in Table 1. No retardate was included who did not have a visual acuity level with Snellen equivalents of 20/29 or better on the Bausch and Lomb Modified Ortho-Rater. It was not considered necessary to administer the Ortho-Rater to normal Ss since they had periodic eye exams and presumably wore corrective lenses where necessary.

One male retardate, who became upset over his inability to find some of the pieces, was replaced by an additional S. With this exception, Ss in

TABLE 1

CHARACTERISTICS OF THE THREE GROUPS

Group	N		$\overline{\text{Age}}$ (in yr.)	SD	$\overline{\text{IQ}}$[a]	SD	$\overline{\text{MA}}$	SD
	M	F						
Retardates	18	14	15.14	1.16	63.63	7.01	9.59	.94
Fourth graders	18	14	9.65	.26	108.75	7.64	10.49	.67
Seventh graders	26	26	12.78	.41	108.62	8.40	13.88	.70

[a] The IQs are from one of the Wechsler intelligence scales for the retarded group, from the Primary Mental Abilities Test for the fourth graders, and from the Hennon-Nelson Test of Mental Ability for the seventh graders. Mean IQs and mean MAs are based on the IQs available on all the retardates, 28 fourth graders, and 50 seventh graders.

all groups appeared to find the task interesting and enjoyable, and all appeared motivated to perform well. Retardates were told prior to testing that if they did well, they would receive a reward (a candy bar or two cigarettes). All the retardates were rewarded upon completing the task. Normals were not materially rewarded because fourth-grade teachers have reported that students who do not participate become upset over not receiving rewards.

RESULTS

Comparison of groups.—Since there were no significant sex, race, or order (blank puzzle first or second) effects within any of the groups, these variables were combined within each group for subsequent analysis.

A two-way analysis of variance on the data shown in Fig. 1 indicated that groups, information level, and their interaction were all significant ($p < .01$ or better). A Duncan multiple-range test indicated that the total search time of the seventh graders was significantly shorter than the total search time of the fourth graders and retardates ($p < .001$), who did not differ from each other.

The most important finding was the significant interaction. As information available in the pieces was reduced, the retardates and fourth graders were more deleteriously affected than were the seventh graders.

Although not necessarily relevant, it might nevertheless be argued that the important Groups × Information Level interaction is merely a function of the ceiling effect imposed by the short latencies on the highest information pieces (Natural Pieces 3, 6, 4, and 7). This cannot be the case with the comparison between the fourth and seventh graders, however, since there were no significant differences on the low-information natural pieces (1, 2, 5, and 8), where no ceiling effect occurs, and large significant differences on the blank pieces. However, the retardates did differ ($p < .01$) from the seventh graders on the low-information natural pieces. Consequently, an additional analysis of variance was performed comparing the retardates and seventh graders on the six lowest information level data points shown on Fig. 1. A significant ($p < .005$) Groups × Information Level interaction again emerged, indicating that the overall significant interaction is not simply a function of ceiling effects imposed by the highest information natural pieces.

Validity of information values.—If the judgments and ratings used to derive information values for the puzzle pieces were valid, search time should have decreased as

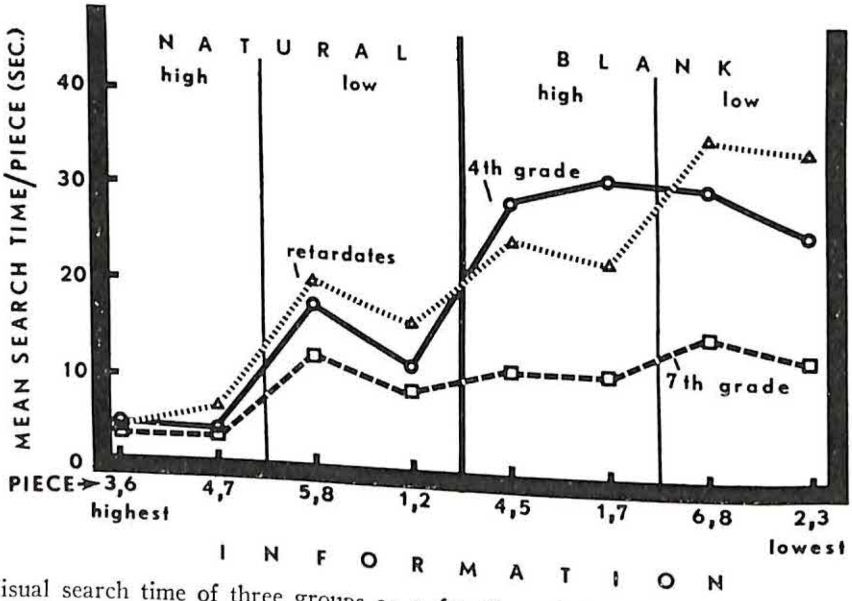


FIG. 1. Visual search time of three groups as a function of decreasing information in the target pieces. (Data points have been halved and the curves smoothed by combining and averaging the search time of each piece with the search time of another piece having identical or adjacent information value.)

TABLE 2

Average Search Time per Piece for Three Groups as a Function of Information Value of Each Piece

Piece No.	Information value	Average search time		
		Retardates	Fourth graders	Seventh graders
		Natural puzzle		
	High			
6	11.00	3.3	3.7	2.1
3	10.29	4.9	5.8	5.0
7	9.00	8.8	5.4	4.0
4	7.43	5.0	3.5	3.4
	Low			
5	3.71	32.5	26.1	16.1
8	2.57	8.0	10.3	9.0
2	2.00	14.2	12.3	7.5
1	1.43	19.0	12.2	11.1
		Blank puzzle		
	High			
4	23	20.1	13.6	7.6
5	20	30.7	45.5	16.4
7	19	24.7	17.4	8.5
1	18	21.9	47.2	15.0
	Low			
6	16	34.7	25.0	13.9
8	14	38.9	38.2	18.4
2	13	16.6	19.1	8.8
3	13	54.0	34.7	18.8

follows: high-information natural pieces < low-information natural pieces; and high-information blank pieces < low-information blank pieces. On a purely subjective basis, one would also predict that search time for low-information natural pieces would be less than for high-information blank pieces since it seems reasonable to presume that on this simple puzzle, even low-information content pieces should be easier to find than blank pieces with high contour information.

Figure 1 shows that, in general, the expected order obtains for all three groups. For each group, the search times on the four pieces at each information level were combined and compared with combined search times (latency) of adjacent information levels. Statistical comparisons (3 levels × 3 groups) were made by t tests for correlated means. Seven of the nine comparisons were significant, all in the predicted direction ($p < .05$ or better, two-tailed tests). The two nonsignificant differences were between low-information natural pieces and high-information blank pieces for the seventh graders and between high- and low-information blank pieces for the fourth graders. The latter result stems from the great difficulty fourth graders had with Pieces 1 and 5, which are located in the upper left quadrant of the puzzle.

Additional findings.—It was of interest to compare the search time for identical puzzle pieces when natural and when blank. As noted, a piece's contour alone should supply less information than its combined contour and content. Consequently, a puzzle piece having low *contour* information when blank and high *content* information when natural should create the greatest difference in search time. This is clearly what happens with the retarded group, as can be seen in Fig. 2 at Points D–D′. The average latency per piece for the combined scores of Pieces 3 and 6 is 40 sec. longer when their contents are eliminated than when they are natural ($p < .001$).

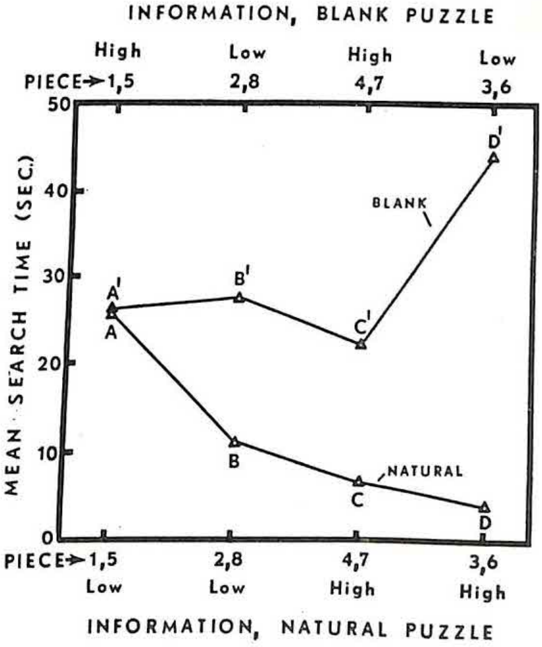


FIG. 2. Retarded group's mean search time when the same puzzle piece is natural and when it is blank. (Each data point is the average search time per piece for two pieces.)

On the other hand, when the contour information of a blank piece is high and its natural content information is low, the smallest difference in search time should occur. In fact, average search time per piece for combined Pieces 1 and 5 when they are natural is not significantly different from when they are blank (see Fig. 2, Points A–A′).

The average search time difference for pieces containing high contour information when blank and high content information when natural, or low information under both conditions, should be intermediate between the two extremes. Differences of 16 and 17 sec. for these conditions bear this out (see Fig. 2, Points B–B′ and C–C′). Differences at B–B′ and C–C′ are significant ($p < .005$), but significantly smaller than the differences at D–D′ ($p < .005$).

Similar results, with the same comparisons statistically significant, were found for the seventh graders, but not for the fourth graders, who performed so poorly on Blank Pieces 1 and 5.

In summary, for the retardates and seventh graders, latency differences for identical pieces depend largely on the content information originally available relative to the contour information remaining in the pieces after the content has been eliminated.

Note also that the relatively smooth curve in Fig. 1 was achieved by combining and averaging results from two pieces with identical or adjacent information values. Results from individual pieces were not nearly as neat (see Table 2). Despite these deviations, however, the finding that seven of the nine comparisons were significant in the expected direction indicates that the information value measures used here were grossly reliable.

DISCUSSION

As previously noted, the present measures of information value proved quite serviceable. However, location appears to influence search latency in this task in certain instances. Note in Table 3 that Piece 5 in both puzzles was relatively difficult for all groups and Blank Piece 2, positioned close to the target piece, was relatively easy in spite of its low-information value. Other factors also play a role. Although Natural Piece 8 was of low-information value relative to the entire picture, the presence of a flower in that puzzle piece probably provided a high measure of independent information, thereby accounting for its short latency. Apparently, both position and independent information value are variables that additionally influence search time, although the major variables are the information in a natural piece relative to the entire puzzle and shape complexity in the blank pieces.

Extreme loss of information resulting from eliminating content is severely and equally debilitating for both the retardates and the fourth graders. It has much less effect on the performance of the seventh graders. Clearly, the seventh graders have additional resources for coping with the loss of information.

In considering the mechanisms used in visual search, Neisser's (1967, p. 94) two-level hypothesis is attractive. Neisser posits an initial preattentive phase during which the organism scans the material for further processing. This is followed by a focal attention phase when the organism concentrates on and analyzes the chosen object. Presumably, poor performance on the present task can be attributed to difficulties in the preattentive phase; for once the organism is cued to the location of the target piece, more intense focal attention is merely confirmatory. Such an analysis applied to the present data suggests that as people mature, they develop additional resources at the preattentive stage of alertness.

The increasing performance differences between high- and low-MA groups have been found previously on other tasks (Siegel, 1968; Spitz et al., 1968). Presumably, the difficulty level could be additionally increased to a point at which all groups fail or function at a chance level. This suggests a general principle that might apply to differences between retardates and normals and between older and younger normals. The principle can be stated as follows: Performance differences of higher and lower MA groups as a function of task difficulty will describe an inverted U-shaped function, with increasing amount of difficulty along the abscissa and increasing amount of difference along the ordinate. In other words, going from ceiling to floor effect, differences between groups will increase to some point and then decrease. Depending on the task, the inverted U-shaped function will be skewed left or right. Depending on the differences in MA, the inverted U-shaped function will flatten or heighten. The advantage of this principle is the reduction of functions derived from many diverse types of experiments to a single comparative function.

REFERENCES

Attneave, F. Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology,* 1957, 53, 221-227.

Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. *Psychological Bulletin,* 1956, 53, 452-471.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Mackworth, N. H., & Morandi, A. J. The gaze selects informative details within pictures. *Perception and Psychophysics,* 1967, 2, 547-552.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Pollack, I., & Spence, D. Subjective pictorial information and visual search. *Perception and Psychophysics,* 1968, 3, 41-44.

Siegel, L. S. The development of the ability to process information. *Journal of Experimental Child Psychology,* 1968, 6, 368-383.

Spitz, H. H. The role of input organization in the learning and memory of mental retardates. In N. R. Ellis (Ed.), *International review of research in mental retardation.* Vol. 2. New York: Academic Press, 1966.

Spitz, H. H. Information transmission in an absolute judgment task with feedback, using normal and retarded subjects. *Journal of Comparative and Physiological Psychology,* 1967, 64, 85-92.

Spitz, H. H., Hoats, D. L., & Holden, E. A., Jr. Numerosity discrimination of tachistoscopically presented dots by mental retardates and normals. *American Journal of Mental Deficiency,* 1968, 73, 127-138.

(Received April 5, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 488–492

PREFERENCE FOLLOWING ADAPTATION OF AN IDEAL AND THE METRIC SCALING OF VISUAL COMPLEXITY[1]

HOBEN THOMAS[2]

Pennsylvania State University

In five experiments, cats ($N = 50$) were adapted to a visually complex stimulus on Trial 1 and then were provided with two different choice stimuli on Trial 2. The *S*s' behavior was assumed to be in correspondence with Coombs' preference choice postulate, and *S*s were therefore expected to select as preferred the choice stimulus nearest to the adaptation stimulus. Two experiments used preference choice data to develop an ordered metric scale of stimulus distances. Three other experiments validated the distance relations obtained; *S*s did select the nearer of two stimuli as predicted.

In a recent study of exploratory behavior (Thomas, 1969), cats wandered among three differentially complex visual stimuli for six trials. On each trial, preference order relations for the stimuli were determined. The more time spent with a stimulus, the more it was said to be preferred. The preference order relations, interpreted with Coombs' (1964) theory of choice as a scaling criterion, indicated that from *S*s' viewpoint, the stimuli could be ordered along a single unidimensional complexity continuum. Having demonstrated behaviorally that a series of stimuli satisfy the conditions for unidimensionality, the problem of determining from behavioral data metric relations among stimuli (i.e., distances between stimuli) is considerably reduced. In fact, given only an order relation among stimuli, some metric relations can already be deduced. To illustrate, let HS, VS, LC, and SC in Fig. 1A represent four ordered stimuli along a single unidimensional continuum. Then $\overline{\text{HS,SC}}$ (where $\overline{\text{HS,SC}}$ means the distance between HS and SC) must be greater than the distance between any other pair of stimuli; HS and SC are end points of the scale. Also, $\overline{\text{HS,LC}}$ must be greater than $\overline{\text{HS,VS}}$ since $\overline{\text{HS,LC}} = \overline{\text{HS,VS}} + \overline{\text{VS,LC}}$; but whether $\overline{\text{HS,VS}}$ is greater or less than $\overline{\text{VS,LC}}$ is an empirical question that cannot be deduced. These examples should demonstrate that some of the distance relations among a set of stimuli can be inferred given an order relation, while others need to be determined empirically.

Determining distances between stimuli from behavior data would seem to be an important first step toward specifying parameters more precisely in studies of, e.g., ex-

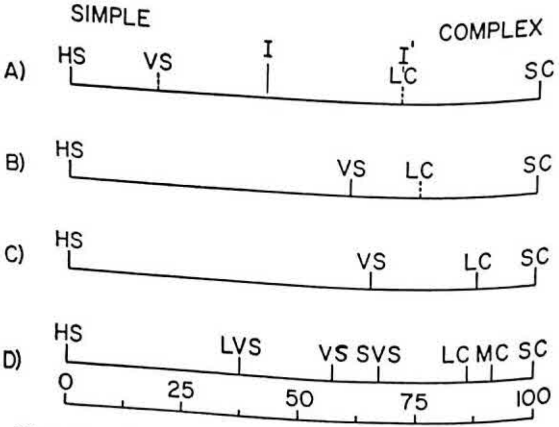


FIG. 1. Stimuli on simplicity–complexity scale. (Stimuli with dotted lines accurate within order-preserving transformation. Distance relations between stimuli with solid lines accurate within monotone transformation. Fig. 1A: *J* scale. I = location prior to adaptation on LC; I′ = location after adaptation on LC. I′ should prefer SC over VS. SC, VS are bilateral to I′; VS, HS are unilateral to I′. Fig. 1B: Relations after Exp. I. Fig. 1C: Relations after Exp. II. Fig. 1D: Distances on interval scale, $\Delta = 1$, and stimuli for Exp. III to V.)

---

[1] This research was supported by Grant MH-04726 from the National Institute of Mental Health, United States Public Health Service. The author wishes to thank J. M. Warren for the use of his laboratory and animals.

[2] Requests for reprints should be sent to Hoben Thomas, Department of Psychology, Pennsylvania State University, 117 Burrowes Building, University Park, Pennsylvania 16802.

ploratory behavior. While scaling procedures can be used without great difficulty in most human research, the problem in animal research is more acute. So far, other than operant procedures, such as Blough's (1958), the problem has apparently been unresearched.

Two of five experiments reported here were designed to determine the metric order relations among four stimuli (i.e., the rank order of the stimulus distances). Three additional studies were designed to test the distance relations obtained in Exp. I and II.

GENERAL METHOD

The procedure for each of the five studies was exactly the same. Only *S*s and stimuli used in each experiment were different.

Subjects.—The *S*s were 50 cats, 10 in each experiment. All *S*s were born and reared in large laboratory cages and averaged 13 mo. of age.

Apparatus.—The main apparatus was a triangular box used in a previous study (Thomas, 1969). Each side was a wooden frame, 6 ft. long × 22 in. high, in which clear glass was mounted. Partitions bisected each interior angle of the box, dividing it into three compartments. The ends of the partitions were angled so as to leave a neutral center portion of the box through which the animals had to pass to enter any compartment, but from which it was difficult to see into any compartment.

The top of the box was covered with translucent plastic. Lamps centered over the top of each compartment provided a bright diffused light for the box interior and for the stimuli, which were attached to the glass walls on the outside of the box. Except for the translucent top and glass walls, the box was painted black.

A rectangular retaining box, 29 × 18 × 30 in. tall with a translucent top, was used to retain *S*s between trials. One portion of one side was glass, through which a stimulus could be viewed.

Stimuli.—The stimuli were large pieces of cardboard cut to fit the glass walls in each side of the triangular box. Complexity was defined as the number of brightness changes within a stimulus, equivalent to the number of black and white elements in a stimulus. The more elements, the more complex it was. For Exp. I and II, the stimuli, in order of increasing complexity, were 4-in. black and white horizontal stripes (HS) with 5½ elements; 3-in. black and white vertical stripes (VS) with 27 elements; two checkerboard patterns, one with 2-in. black and white squares (LC) with 280 elements, the other with about 5,900 ½-in. squares (SC). These stimuli, all with equal portions of black and white, have been shown to lie on a single unidimensional complexity continuum (Thomas, 1969). The stimuli for Exp. III, IV, and V, similarly constructed, are described subsequently.

Procedure.—The *S*s were run at night. An *S* was placed into the neutral center portion of the triangular box, and the top was lowered into place. The compartment lights were turned on to illuminate the box, and a trial was begun. The *E* observed S from above the box. When an *S* entered a compartment, i.e., when his head crossed an imaginary line formed by the end of a compartment partition and wall of the box, one of three switches was closed, activating a pen on an event recorder (Gerbrands). Thus, length of time in each compartment was recorded.

Each *S* was run for two consecutive 10-min. trials. On Trial 1, the adaptation trial, identical "adaptation" stimuli were in each compartment. When the trial ended, *S* was placed in the retaining box, which had secured to its glass wall half of an adaptation stimulus. Two of the three identical adaptation stimuli in the triangular box were randomly selected and were replaced with two different "choice" stimuli. The third compartment retained the adaptation stimulus. The *S* was again placed in the triangular box, and a 10-min. choice trial was run.

RATIONALE

In Coombs' (1964) unfolding theory of preference, both individuals and stimuli may be ordered along a common scale called a *J*, or joint, scale. In this case, the *J* scale is a complexity scale. An individual's location on the scale is called his ideal. The Coombs' choice postulate states that the "individual's preference ordering of the stimuli from most to least preferred corresponds to the rank order of the absolute distances of the stimulus points from the ideal point, the nearest being most preferred [p. 80]." It is assumed that *S*'s choice behavior is in correspondence with the choice postulate, and further, that the adaptation trial causes his ideal to coincide with the adaptation stimulus' point on the scale. Given these assumptions, *S* should prefer, of the choice stimuli, the stimulus closest to the adaptation stimulus. Interest in these studies is focused on preference between the two choice stimuli. Whether *S* prefers or avoids (relative to the choice stimuli) the adaptation stimulus present on the choice trial is of no particular interest except as it relates to the adaptation assumption to be discussed.

Only choice stimuli bilateral to the individual's ideal point (see Fig. 1A), i.e., one choice stimulus on each side of the ideal, have metric implications; unilateral choice stimuli (both stimuli on the same side of the ideal) provide information on order relations between the stimuli, but not information on metric relations. In all the present experiments the choice stimuli are bilateral to *S*'s adapted ideal.

TABLE 1

TIME WITH STIMULI ON CHOICE TRIALS

Exp.	*S*	Relative complexity level		
		Low	Medium	High
I		HS	VS	SC
	1	211	115	107
	2	85	167	241
	3	27	135	34
	4	124	115	26
	5	164	115	165
	6	50	229	288
	7	130	13	385
	8	67	10	186
	9	10	5	313
	10	115	156	284
II		VS	LC	SC
	1	30	15	278
	2	123	246	211
	3	85	341	154
	4	261	49	212
	5	258	1	195
	6	212	69	304
	7	33	471	81
	8	84	190	243
	9	7	33	458
	10	124	15	203
III		HS	LVS	SC
	1	123	389	69
	2	110	126	308
	3	54	238	49
	4	488	62	6
	5	287	15	163
	6	90	48	109
	7	154	209	173
	8	485	68	23
	9	331	108	114
	10	233	84	108
IV		LC	MC	SC
	1	269	44	183
	2	103	201	248
	3	121	328	72
	4	183	201	164
	5	89	246	64
	6	285	131	143
	7	209	188	129
	8	238	171	138
	9	185	171	189
	10	293	82	168
V		VS	SVS	MC
	1	49	165	194
	2	197	87	153
	3	167	25	67
	4	176	115	187
	5	194	241	99
	6	172	233	88
	7	281	72	86
	8	138	70	47
	9	156	150	86
	10	177	86	121

Note.—All times in seconds. Time in neutral center of box = 600 − low − medium − high.

EXPERIMENT I

Stimuli.—The VS was the adaptation stimulus; HS and SC were the choice stimuli.

Results and discussion.—In Table 1 the total times spent by each *S* in the stimulus compartments on the choice trials for Exp. I and the remaining experiments are reported. The Wilcoxen matched-samples test (Hays, 1963, p. 635) was applied to the time score preferences for HS and SC for each *S*. Eight *S*s preferred SC over HS, i.e., spent more time in the compartment with SC than in the compartment with HS; and *S*s spent significantly more time with SC than with HS, $T = 7$, $p < .05$ (two-tailed).

Given this finding, assuming the choice postulate holds, it may be concluded that SC is closer to VS than HS is to VS (see Fig. 1B). Thus, it is possible to order the following distance relations: $\overline{\mathrm{HS,SC}} > \overline{\mathrm{HS,LC}} > \overline{\mathrm{HS,VS}} > \overline{\mathrm{VS,SC}}$, leaving the relations between $\overline{\mathrm{VS,LC}}$ and $\overline{\mathrm{LC,SC}}$ to be determined (again, see Fig. 1B). The distance $\overline{\mathrm{HS,LC}}$ can be inferred to be $> \overline{\mathrm{HS,VS}}$, given the finding that VS is closer to SC than to HS.

EXPERIMENT II

Stimuli.—The LC was the adaptation stimulus; VS and SC were choice stimuli.

Results and discussion.—Eight of 10 *S*s preferred SC over VS, and significantly more time was spent with the SC stimulus, $T = 5$, $p = .02$ (two-tailed).

It is now possible to order all pairs of distances: $\overline{\mathrm{HS,SC}} > \overline{\mathrm{HS,LC}} > \overline{\mathrm{HS,VS}} > \overline{\mathrm{VS,SC}} > \overline{\mathrm{VS,LC}} > \overline{\mathrm{LC,SC}}$, a diminishing ordered metric scale where the distances between adjacent pairs become progressively smaller. The situation is illustrated in Fig. 1C.

These metric relations were converted to an interval scale using Goode's procedure under the constant Δ assumption (Coombs, 1964). Scale values did not change greatly for various values of Δ (except for $\Delta = 0$, which caused LC and SC to coincide). The value $\Delta = 1$ was selected as reasonable, the scale was limited from 0 to 100, and the obtained values were as follows: HS = 0, VS = 57.1, LC = 85.7, SC = 100. These stimulus scale values were plotted against the number of black and white elements in each stimulus, and a curve was fitted by eye to the four points. By reference to the curve, stimuli could be selected with specified numbers of black and white elements and specified stimulus scale values.

EXPERIMENT III

It has been shown that with exposure to complex visual stimuli, *S*s tend to select increasingly complex stimuli (e.g., Dember, Earl, & Paradise, 1957; Thomas, 1969). Therefore, it was reasoned that the most demanding test of the metric scale relations would be to predict that *S*s should prefer, as the nearest bilateral choice stimulus, the stimulus of least complexity. Additionally, if *S*s move along a complexity scale toward increasing complexity, as has been demonstrated (Thomas, 1969), they would be moving *away* from the simple stimulus toward the complex stimulus.

Stimuli.—New adaptation stimuli were made with large 5-in. vertical stripes (LVS), each with 11 elements. The corresponding scale value of 37 falls in the lower central portion of the stimulus scale, as Fig. ID illustrates. Choice stimuli were HS and SC. It was predicted that HS would be preferred over SC.

Results.—Seven *S*s preferred HS over SC, and more time was spent with HS than SC, $T = 12$, $z = 1.59$, $p < .06$ (one-tailed).

EXPERIMENT IV

Stimuli.—New adaptation stimuli were made with medium-size 1.25-in. checkers (MC), each with 616 elements. The corresponding scale value of 91 falls in the upper portion of the stimulus scale, as Fig. 1D illustrates. Choice stimuli were LC and SC. It was predicted LC would be preferred over SC.

Results.—Eight *S*s preferred LC over SC, and more time was spent with LC than with SC, $T = 11$, $z = 1.69$, $p < .05$, (one-tailed).

EXPERIMENT V

Stimuli.—New adaptation stimuli were made with small 2-in. vertical stripes (SVS), each with 28 elements. The corresponding scale value of 67 falls in the upper central portion of the stimulus scale (see Fig. 1D). Choice stimuli were VS and MC; it was predicted VS would be preferred over MC.

Results.—Eight *S*s preferred VS over MC, and more time was spent with VS than with MC, $T = 10$, $z = 1.78$, $p < .04$ (one-tailed).

GENERAL DISCUSSION

Experiments I and II demonstrated that it is possible to generate metric relations among stimuli from behavioral data. Experiments III, IV, and V demonstrated that the metric relations obtained hold for an independently generated set of adaptation stimuli.

Although the distances obtained were expressed on an interval scale, the tests of the obtained distance relations cannot be said to be tests of the interval scale. Scale separations between the adaptation stimulus and choice stimuli predicted as being preferred were, on the average, about half as great as the distance between the adaptation stimulus and the more complex choice stimulus. Thus, the tests of distance relations in Exp. III, IV, and V are best perceived as being tests of an ordered metric scale. Future studies designed to test preferences for choice stimuli more similar in proximity to the adaptation stimuli will undoubtedly have to treat preference behavior as a stochastic process. This topic was discussed by Coombs (1964).

There is no certainty that the procedures used here would always produce a completely ordered metric scale. More likely to result are partially ordered metrics (a situation in which some pairs of distances have tied ranks), particularly if many stimuli are used. Further, the strategy for determining distances by this method is an exercise in *E* judgment. For example, if in Exp. I, HS had been preferred over SC, fewer metric relations could have been inferred, and more than two experiments would have been required to obtain an ordering of stimulus distances. Too, the procedure becomes extremely cumbersome as the number of stimuli increases.

Probably different scaling techniques will lead to different metrics; in fact, this would be expected from theory. The stimulus scale developed here is a preference scale. Assessing stimulus distances from similarity data (e.g., Dember, 1957) may lead to different metrics. Similarity and preference data are believed to represent quite different psychological processes (Coombs, 1964).

Whether the adaptation trial caused *S*s' ideals to coincide with the adaptation stimulus point is an important question, but it is difficult to answer directly. Demonstrating that the adaptation trial altered *S*s' responses toward the adaptation stimulus on the choice trial would be indirect but partially confirming evidence. In an earlier experiment (Thomas, 1969), the first choice trial was similar to the choice trials here, except there was no previous adaptation trial. In that study, 12 of 21 cats preferred most, on their first choice trial, the "adaptation stimulus" (i.e., the stimulus of intermediate complexity). In the experiments reported here, the same proportion of *S*s might be expected to prefer most the adaptation stimulus if the adaptation trial had failed to alter *S*s' behavior. In fact, only 12 of 50 *S*s preferred *most* the adaptation stimulus. The probability of observing 12 in a sample of 50 from a binomial population with $p = 12/21$ is small, $p < .001$ (two-tailed). Thus, it appears that *S*s were adapted to the stimuli, or at least their behavior toward the adaptation stimuli was altered from what might be expected for unadapted *S*s.

Incidentally, the finding that *S*s avoid the adaptation stimulus on the choice trial, preferring a stimulus slightly discrepant from adaptation level, is in agreement with the McClelland discrepancy from affect hypothesis (Haber, 1958; McClelland & Clark, 1953). Also, it is quite probable that under some conditions, preference for a choice stimulus would increase with increasing distance between the adaptation and choice stimuli. This possibility needs to be explored.

These results should make it possible to specify more quantitatively parameters in some theories of motivation. For example, Dember and Earl (1957) predicted that *S*s would be motivated by stimuli slightly above their complexity ideals. How far from the ideal point stimuli can be and still be motivating has never been determined, but this should now be possible.

Finally, these procedures for obtaining metric relations might well be employed in studies similar to Haber's (1958) classic experiment where there is no obvious psychologically relevant physical stimulus measure.

REFERENCES

Blough, D. S. A method for obtaining psychophysical thresholds from the pigeon. *Journal of the Experimental Analysis of Behavior,* 1958, **1**, 31–43.

Coombs, C. H. *A theory of data.* New York: Wiley, 1954.

Dember, W. N. The relation of decision-time to stimulus similarity. *Journal of Experimental Psychology,* 1957, **53**, 68–72.

Dember, W. N., & Earl, R. W. Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review,* 1957, **64**, 91–96.

Dember, W. N., Earl, R. W., & Paradise, N. Response by rats to differential stimulus complexity. *Journal of Comparative and Physiological Psychology,* 1957, **50**, 514–518.

Haber, R. N. Discrepancy from adaptation level as a source of affect. *Journal of Experimental Psychology,* 1958, **56**, 370–375.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

McClelland, D. C., & Clark, R. A. Antecedent conditions for affective arousal. In D. C. McClelland, J. W. Atkinson, R. A. Clark, & E. L. Lowell (Eds.), *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

Thomas, H. Unidirectional changes in preference for increasing visual complexity in the cat. *Journal of Comparative and Physiological Psychology,* 1969, **68**, 296–302.

(Received April 7, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 493–498

UNRESTRICTED CLASSIFICATION BEHAVIOR AND LEARNING OF IMPOSED CLASSIFICATIONS IN CLOSED, EXHAUSTIVE STIMULUS SETS[1]

BERT ZIPPEL[2]

Hunter College of the City University of New York

The *S*s were required to classify compact and distributed stimulus sets according to partitions varying in frequency, as offered by *S*s in a previous free classification study. In Exp. I, performance with compact sets was predictable from knowledge of preferred classifications, although a single-unit S–R theory could also predict the differences. In Exp. II, partitions of compact and distributed sets were used. Imposed partitions corresponding with preferred classifications produced less errors than when partitions did not so correspond. The data of both studies indicate that the interaction of different partitions with different forms of simulus sets produce codable patterns that vary greatly in complexity.

Increasing attention has been devoted to stimulus organization in human learning (Mandler, 1967a, 1967b). These studies have indicated that *S*s tend to organize any set of stimulus materials, as evidenced particularly in free recall techniques of assessing the learning of lists of stimulus forms. Fallon and Battig (1964) have indicated that *S*s' preferences for organizing stimulus materials influence later imposed learnings. Behavior in such classification and learning studies could be dependent on the properties of stimulus sets. One property is whether the stimulus sets are compact or distributed (Shepard, Hovland, & Jenkins, 1961). A compact stimulus is a single unit that embodies all of the information presented, whereas a distributed stimulus requires separate units presented at the same time to indicate the information conveyed by the set member. Studies using this distinction have been reported by Shepard et al. (1961) and Zippel (1968). Also, Chang and Shepard (1964) tried to extend the distinction from geometric stimuli and pictorially meaningful objects to verbal materials, with only moderate success.

In general, all organisms come to a stimulus situation with previous patterns of response to portions of the situation. Such response patterns could be thought of as operant modes of element clustering (OMEC) behaviors, and these OMEC behaviors may or may not be stable or consistent with required response patterns in any given situation.

Few studies have systematically ascertained human *S*s' OMEC behaviors in relation to the demands of imposed learning tasks. The purpose of the present studies is to relate OMEC behavior with different partitions of compact and distributed stimulus sets. In both of the studies, the compact and distributed sets are closed (Cohen, 1966) and exhaustive (Esper, 1925). The first study compares the rate of learning various partitions of compact stimulus sets where the partitions vary in frequency of occurrence as OMEC behaviors.

EXPERIMENT I

Shepard et al. (1961) determined that there are six basic partitions of three-attribute, two-values-per-attribute stimuli into two groups of four set members each. Zippel (1968) generated a number of stimulus sets and instructed *S*s to group their members in any way they wished. Most of the *S*s sorted the cards into four groups of two members each, using two attributes for the basis of their sorting. As a result of these data, the

---

[1] This research was supported by the graduate division of the City University of New York.

[2] Requests for reprints should be sent to Bert Zippel, Department of Psychology, Hunter College of the City University of New York, 695 Park Avenue, New York, New York 10021.

different partition rules possible for classifying such sets into four groups of two cards each were determined. The 6 partition rules of Shepard et al. and the 12 rules determined by this *E*, along with the number of different ways each rule could be applied for a given set of attributes, are described in Table 1. The letters A, B, and C represent arbitrary designations of the attributes of a set. The uppercase or lowercase designations indicate which of the two values of the attribute is being considered. Plus signs (+) are used to separate categories, while $\wedge$ signs indicate conjunction within a category. Thus, e.g., *Z6* expanded would give the categories AB, Ab, $aBC \wedge abc$, and $aBc \wedge abC$. With this partition rule, there would be 12 possible arrangements of a set that would fit the description. That is, if size were one of the attributes, then size could be A in four of the arrangements or B or C in the remaining arrangements. Table 1 also presents the number of *S*s who partitioned the sets in the earlier classification study (Zippel, 1968) according to these and other partition rules.

From the 18 *SHJ* and *Z* partition rules, a subset of five was selected for imposed classification learning with compact sets: *Z1, Z2, Z12, SHJ1*, and *SHJ6*. It was hypothesized that in a complete learning situation (Haygood & Bourne, 1965), *S*s would learn to classify a set of stimuli partitioned according to rules that are consistent with frequent OMEC responses (*SHJ1* and *Z1*) faster than sets partitioned according to partition rules that are less often chosen (*Z2, Z12, SHJ6*) in an OMEC situation. A sixth condition was added to the study to determine the effect of learning a separate response to each of the members of a set.

Method

Subjects.—Six volunteer college *S*s were randomly assigned to each of the six learning conditions for a total of 36 *S*s. None of the *S*s had previous experience with an experiment involving the kinds of stimulus or response materials used in the present study.

Construction of the stimulus displays.—Stimulus sets were generated which were composed of geometric forms of 3 attributes with two values per attribute. An attribute is defined as a label commonly assigned to a set of objects that

TABLE 1

CLASSIFICATIONS OF THREE ATTRIBUTE STIMULUS SETS INTO TWO GROUPS OF FOUR AND FOUR GROUPS OF TWO

Partition	Algebraic description	Groupings	OMEC frequency
SHJ1	$A + a$	3	3
SHJ2	$AB \wedge ab + Ab \wedge aB$	3	
SHJ3	$AB \wedge aC + Ab \wedge ac$	12	
SHJ4	$AB \wedge C(Ab \wedge aB) + ab \wedge c(Ab \wedge aB)$	4	
SHJ5	$AB \wedge a(BC \wedge bc) + Ab \wedge a(Bc \wedge bC)$	12	
SHJ6	$A(BC \wedge bc) \wedge a(Bc \wedge bC) + A(Bc \wedge bC) \wedge a(BC \wedge bc)$	1	(1)[a]
Z1	$(A + a)(B + b)$	3	15
Z2	$A(B + b) + a(C + c)$	6	1
Z3	$AB + ab + (C + c)(Ab \wedge aB)$	6	
Z4	$(A + a)[(Bc \wedge bC) + (BC \wedge bc)]$	3	
Z5	$AB + aC + bc + AbC \wedge aBc$	8	
Z6	$A(B + b) + a(BC \wedge bc + Bc \wedge bC)$	12	
Z7	$AB + bC + a(BC \wedge bc) + c(Ab \wedge aB)$	24	
Z8	$AB + ab + AbC \wedge aBc + Abc \wedge aBC$	6	
Z9	$(A + a)(Bc \wedge bC) + ABC \wedge abc + Abc \wedge aBC$	6	(2)
Z10	$AB + a(Bc \wedge bC) + b(AC \wedge ac) + Abc \wedge aBC$	24	
Z11	$A(bC \wedge Bc) + B(AC \wedge ac) + b(Ac \wedge aC) + a(bc \wedge BC)$	6	
Z12	$ABC \wedge abc + ABc \wedge abC + AbC \wedge aBc + Abc \wedge aBC$	1	(4)
FD1	$3 \geqslant x_1 \geqslant 0$	4	(9)
	Other	—	1 (6)

[a] Parentheses indicate frequency of OMEC groupings with distributed sets. No parentheses indicate frequency of groupings with compact stimulus sets.

can serve as a focus for further subdivisions of a set, such as color or shape. Attributes are distinguished from dimensions which imply subdivision of an attribute according to orthogonal qualities that vary in magnitude. For example, color is an attribute that is composed of the dimensions of hue, brightness, and saturation, whereas the attribute of shape cannot be similarly dimensionalized (see Torgerson, 1958, pp. 247–250 for a similar use and Slamecka, 1968, for a different use of terminology). A list of 11 attributes was created, from which sets of 3 attributes could be drawn at random. The attributes were the size, color, shape, thickness, solidness, and location of the figure, and the number, thickness, solidness, slant, and color of lines drawn inside the figure. Various specific sets of stimuli could then be generated by randomly varying a subset of these attributes and randomly fixing the values of the remaining attributes of the pool.

Design.—Three different sets of three-attribute, two-values-per-attribute compact stimuli were generated. Three random orders of each set were constructed, and the figures drawn on 3 × 5 index cards were inserted into plastic file card holders, each of which then contained all 24 cards of the three replications of a given set. A second and third set of such random orders was constructed for each different stimulus set. Which of the random orders of the sets was presented to a given *S* was randomly determined. All *S*s were exposed to the task with each of the three different sets in random order.

The *S* went through the card file as many times as was necessary to reach a criterion of either 32 errorless presentations or 168 presentations, whichever came first. Each *S* had the same type of partition to learn on three successive tasks. The names for the classes were single letters of the alphabet. These names were randomly determined for each *S* for each set and presented to *S* on a separate card so that no memory was involved in learning the group names. The particular arrangement for a partition rule was also randomly determined for each *S* since some of the partition rules had more than one possible grouping of attributes.

Results and Discussion

Many *S*s did not reach the criterion of 32 errorless presentations, even with the third task. The percentage of performances reaching the criterion of 32 errorless presentations across the three successive tasks ranged from 78% for *SHJ1* to 6% for *SHJ6*, with a median of 33% among the six different partitions. Thus the error data were analyzed. There is a problem in attempting to compare directly the results for the *SHJ* partitions, the *Z* partitions, and the paired-associates task in that the probability of correct guessing is .50 for *SHJ*, .25 for *Z*, and .125 for the paired-associates task. An error measure was taken that represented the proportionate difference between total errors made and chance errors expected for each *S* on each task. Figure 1 presents the mean proportionate errors for each group in each of the three successive tasks. An analysis of variance for this two-factor experiment with repeated measures on the task factor (Winer, 1962) showed no significant differences due to first, second, and third task, $F\ (2,\ 60) = 1.47,\ p < .25$. There was no significant interaction among tasks and the different partitionings, $F\ (10,\ 60) = .08$. There was a significant difference in performance as a result of the various partitions, $F\ (5,\ 30) = 12.8,\ p < .01$. The Newman-Keuls method of testing the difference between all possible pairs of means (Winer, 1962) showed the *SHJ6* and *Z12* partition performances to be significantly different

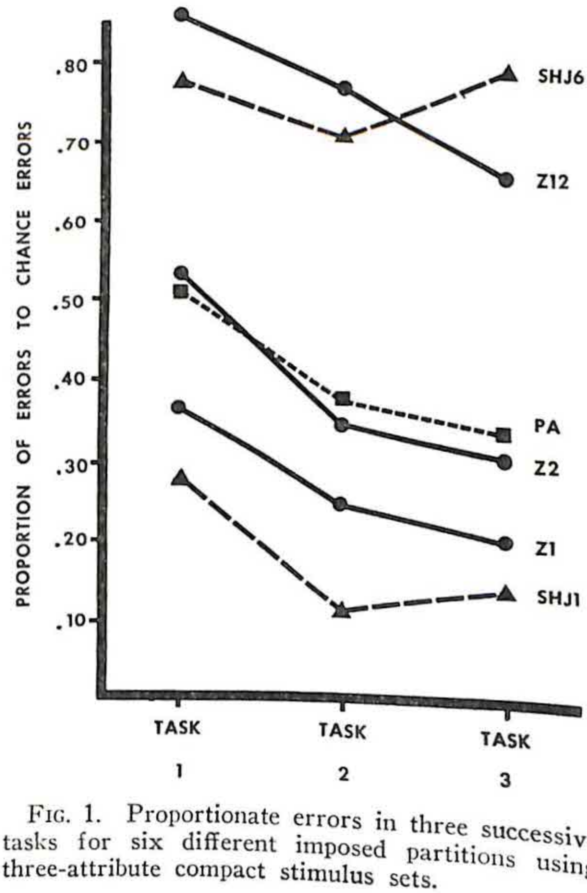


FIG. 1. Proportionate errors in three successive tasks for six different imposed partitions using three-attribute compact stimulus sets.

from all other partition performances at least at the .05 level of significance.

As the control paired-associates task performance is closer to *Z1, Z2*, and *SHJ1* partition performance than the more difficult *Z12* and *SHJ6* partition performance, there is no support for a hypothesis that ease of learning is a direct function of the number of response alternatives available. From the obtained data it also seems to make little difference whether the responses are consistently tied to values of the same two attributes as in *Z1* or whether there is a shift in relevant attributes from group to group as in *Z2*. What seems of central importance among these two partitions is that there is a consistent pairing of two values for each response.

The sets partitioned according to *SHJ6* and *Z12*, which proved to be most difficult to learn, differ from the other conditions in that no conjunctive combination of values is uniquely associated with any given response. The order of difficulty apparent from Fig. 1 would seem to support the hypothesis that difficulty is a function of the number of values that are combined conjunctively to a response which is consistent with studies of learning conjunctive concepts (Bourne, 1966).

The learning data of the present study indicate that frequency of OMEC behaviors help to make only gross predictions as to order of learning sets classified according to imposed partition rules. An explanation involving response competition seems as able to explain the findings of learning difficulty as any hypothesis testing model, especially in light of the overall low degree of learning in the various conditions. Thus, the present study offers only tenuous support to the notion that learning sets classified according to imposed partitions is related to OMEC behavior. The following study reports another attempt to demonstrate a relationship between frequency of OMEC behavior and imposed classification learning.

EXPERIMENT II

Shepard et al. (1961) found that, with the exception of *SHJ1*, distributed sets partitioned according to the *SHJ* rules were more difficult to classify than compact sets. In the OMEC study, Zippel (1968) found that distributed sets are classified differently from compact sets, as indicated in Table 1. Thus, the difficulty of the distributed sets found by Shepard et al. may have been due to requests to classify according to partitions that *S*s would more readily apply to compact sets. Indeed, Fallon and Battig (1964) found that *S*s preferred to classify a three-attribute, two-values-per-attribute distributed set in a way equivalent to a *Z8* partition rule rather than to a *Z1* partition rule, and that learning to classify according to a *Z8* partition rule proceeded faster than classifying according to a *Z1* partition rule. In the classification tasks employed in the OMEC study, three-attribute distributed sets were classed most often by what might be called a frequency distribution rule, which grouped the stimuli into subsets according to the number of occurrences of the value that varied from position to position (*FD1*). For example, if the attribute of size was varied, an *S* would make groups of from zero to three small figures. It was hypothesized, therefore, that learning to classify a set of compact stimuli partitioned according to a *Z1* partition would be faster than learning to partition a distributed set according to this rule. Further, a set partitioned according to an *FD1* rule would be learned faster for distributed than for compact sets. Paired-associates tasks were included for both distributed and compact sets. A complicating factor with the application of the *FD1* partition to distributed sets is that the rule is confounded with the perceptual characteristics of the set. Therefore, another control condition was introduced in which the rule conflicted with the perceptual nature of the set members. This was accomplished by randomly choosing one member of the set other than one where all elements were the same (such as two small and one large, rather than three small) and forming the other groups according to the number of differences from this nonperceptual anchor.

Method

Subjects.—The data from *S*s in the previous study from the *Z1* and control groups were used for this analysis. Six more college volunteer *S*s were assigned to each of five new groups: *FD1* with compact sets; and with distributed sets, *Z1*, *FD1*, paired-associates, and *FD1* with arbitrary anchor.

Construction of the stimulus displays.—The same compact sets used in Exp. I were used in this study, and the distributed sets were constructed from the same pool of attributes by randomly varying 1 of the 11 attributes and randomly fixing the values of the remaining 10 attributes.

Design.—The sets were organized and administered to *S*s as described in Exp. I. For those groups where the *FD1* partition was applied, *S*s were given the four letters that would serve as responses, but were further instructed that they would be grouping the pictures into two groups of one and two groups of three members each. They were not told which response letters were associated with the single-member or triple-member groups.

Results and Discussion

In this study, also, most *S*s did not reach the criterion of 32 errorless presentations. The percentage of *S*s reaching that criterion for the various conditions across the three tasks ranged from 78% for *FD1* applied to distributed sets to 6% for *FD1* applied to compact sets with a median of 39%. Thus, again, the analysis of errors was undertaken. The comparison of the errors in each of the groups offers difficulty in this study, as the *FD1* partition introduces the possibility of probability matching in *S*s responding. To determine the extent of any probability matching, the frequency of error responses for each response was determined for conditions in which four response letters were used. The results showed no differences in trend among the conditions. Nor was there any increase in the responses appropriate for three stimulus set members from the first through the third task except where the *FD1* partition was applied to compact stimulus sets. In this condition the two responses which actually accounted for 75% of the labeling were chosen 55%, 65%, and 66% of the time in the first, second, and third tasks, respectively. However, overall, a model assuming that each of the four responses is used equally often seems most appropriate for the data. Thus in four response alternative conditions, it was assumed that the probability of a correct response was .25, as compared with .125 for eight response alternative conditions. Figure 2 presents the mean proportionate errors for the groups in the three successive tasks. Considering a particular partition as appropriate or inappropriate to a set or a paired-associate task, a three-factor analysis of variance was constructed, with the first factor being form of partition, the second factor form of stimulus set, and the third factor the performance on the three successive tasks with repeated measures on the task factor (Winer, 1962). The nonperceptual anchor group was not analyzed in this comparison.

There was a nonsignificant trend for performance with the distributed stimulus sets to be better than with the compact sets, $F(1, 30) = 2.61$, $p > .10$. In this study the improvement from task to task was significant, $F(2, 60) = 15.41$, $p < .01$, as was the main hypothesis under test indicating that where partitions were appropriate, there were less errors than in tasks where partitions were not appropriate to the sets, $F(2, 30) = 9.02$, $p < .01$. No interactions approached significance. Again, the paired-associates learning tasks did not prove to be

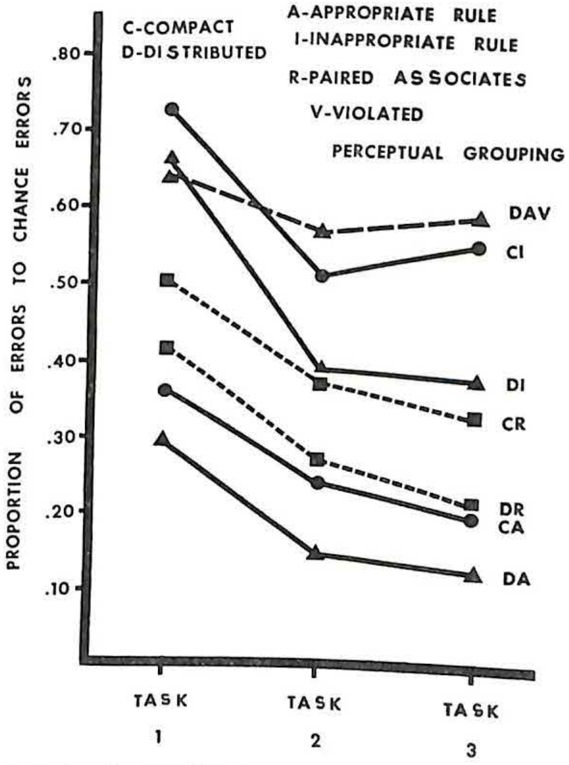


FIG. 2. Proportionate errors in three successive tasks for seven different imposed partitions using three-attribute compact and distributed sets.

much more difficult than the sets with appropriate partitions. A separate analysis to differentiate between the distributed sets using the *Z1* partition and the *FD1* partition with no perceptual anchor was not significant. This would indicate that a distributed set partitioned according to a frequency distribution rule is easy to learn only when the rule coincides with perceptual characteristics of the stimulus set.

In Exp. II the influence of OMEC behavior on classification performance is more evident than in Exp. 1. The preferred OMEC pattern when in agreement with the imposed partition offers better performance than when the partition to be learned is in conflict with the preferred OMEC behavior. In looking at the data of individual *S*s, no consistent patterns were observed which would indicate that *S*s learned the group labels sequentially, and where this seems to be the case for an *S* on a given task, there is no evidence that this continues to be the case from task to task.

The procedures used in these studies do not maximally facilitate the actual learning or verbalization of partition rules. For example, a *Z12* partition rule could be verbalized as follows: "The small black triangle and its opposite is an N, the small black circle and its opposite is a P," etc. The lack of significant improvement from first to third task performance in Exp. I and the limited improvement shown from task to task in Exp. II indicate that *S*s did not, in general, develop a conception of partition rule structures, but rather were operating primarily on the basis of single-unit S–R formations (Kendler & Kendler, 1968). The data of both studies indicate that the interaction of different partition rules with different forms of stimulus sets produces codable patterns that vary greatly in complexity.

It was noted earlier that investigators have not been concerned with carefully differentiating stimulus properties, and these distinctions do elicit different behaviors from *S*s. Thus, the results of these studies should not be generalized beyond the types of stimuli used. Further, the kinds of distributed stimuli used in the present study differ from those initially developed by Shepard et al. (1961). Other OMEC studies are underway to determine whether different forms of generating distributed stimuli make a difference.

REFERENCES

Bourne, L. E., Jr. *Human conceptual behavior.* Boston: Allyn & Bacon, 1966.

Chang, J.-J., & Shepard, R. N. Meaningfulness in classification learning with pronounceable trigrams. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3**, 85–90.

Cohen, B. H. Some-or-none characteristics of coding behavior. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 182–187.

Esper, E. A. A technique for the experimental investigation of associative interference in artificial linguistic material. *Language Monographs,* 1925, **1**, 1–47.

Fallon, D., & Battig, W. F. Role of difficulty in rote and concept learning. *Journal of Experimental Psychology,* 1964, **68**, 85–88.

Haygood, R. C., & Bourne, L. E., Jr. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review,* 1965, **72**, 175–195.

Kendler, H. H., & Kendler, T. S. Mediation and conceptual behavior. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967. (a)

Mandler, G. Verbal learning. In, *New directions in psychology.* Vol. 3. New York: Holt, 1967. (b)

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. Learning and memorization of classifications. *Psychological Monographs,* 1961, **75**(13, Whole No. 517).

Slamecka, N. J. A methodological analysis of shift paradigms in human discrimination learning. *Psychological Bulletin,* 1968, **69**, 423–438.

Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Zippel, B. Free classification of compact and distributed stimulus sets. *Proceedings of the 76th Annual Convention of the American Psychological Association,* 1968, **3**, 45–46.

(Received April 10, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 499–502

OBSERVING RESPONSES, ATTENTION, AND THE OVERTRAINING REVERSAL EFFECT[1]

PETER D. EIMAS[2]

Brown University

Third-grade children were trained on a successive discrimination and its reversal. The apparatus permitted direct and independent measurement of observing responses and hence a finer analysis of the course of reversal learning. The *S*-controlled observing responses functioned to select a single stimulus value from the relevant or irrelevant dimension to which *S* then made an instrumental response. Discrimination learning required acquisition of both the correct observing and instrumental responses. Upon attaining criterion on the original discriminaton, half of the *S*s were reversed immediately, while the remaining half received 25 overtraining trials and then reversal training. Overtraining facilitated reversal learning and, moreover, did so in a manner that confirmed predictions from attentional models of discriminative learning; overtraining, despite producing a small increment in perseverative errors, decreased the number of midreversal errors by reducing the number of observing responses to the irrelevant dimension.

Many studies with both animals and young children have found that discrimination reversal learning is facilitated by overtraining on the original discrimination (for reviews of this literature see Eimas, in press; Mackintosh, 1965; Wolff, 1967). Although the overtraining reversal effect (ORE) has by no means occurred in all studies, it has generally been obtained in animals with difficult discriminations and at times with easy discriminations when large rewards were used. Similarly, in children the ORE has occurred most frequently with difficult discriminations.

A variety of possible mechanisms has been offered in attempts to explain the ORE. For example, Reid (1953) hypothesized that overtraining increases the probability of an organism developing a general "response of discriminating." D'Amato and Jagoda (1961) related the effect to a diminution of avoidance tendencies to the negative stimulus, which was assumed to occur during overtraining. Capaldi and Stevenson (1957) suggested that overtraining increases the discriminability of the situational change that occurs when the reinforcement contingencies are suddenly reversed. Finally, several theorists (Lovejoy, 1966; Sutherland, 1964; Zeaman & House, 1963) have assumed that overtraining strengthens attention to the relevant dimension and in this manner facilitates reversal learning.

With specific regard to the latter mechanism, attentional models assume that discriminative learning requires the acquisition of a chain of two responses. The first is an attentional response to the relevant dimension and the second is an instrumental response of choice to one of the outputs of the attentional response: the positive stimulus value. A further assumption is that both responses may, under appropriate conditions, transfer across problems. Thus in a reversal shift, particularly when there is no change in the stimulus arrangements, positive transfer of the attentional response and negative transfer of the instrumental response would occur. In the mathematical explication of these models (Lovejoy, 1966; Zeaman & House, 1963), it has been demonstrated that with a difficult original discrimination, the probability of the correct instrumental response more closely approximates unity at criterion than does the attentional response. Hence, overtraining is expected to produce a relatively large increment in the probability of making the correct attentional response as compared with the increment in the probability of making the correct instrumental response. From these assumptions, the prediction is for reversal learning after overtraining to be marked by a number of perseverative errors followed by acquisition of the new instrumental response. Without overtraining, reversal learning should be characterized by fewer perseverative errors, but by responses to irrelevant dimensions. The latter results from a further lowering of the probability of attending to the relevant dimension after several extinction (error) trials and, consequently, an increase in the probability of attending to irrelevant dimensions. After these events, reacquisition of the relevant attentional response and acquisition of the new instrumental response are predicted to occur. Overtraining, in essence, is predicted to eliminate or at least greatly reduce the period of responding to irrelevant dimensions, thereby facilitating reversal learning.

---

[1] This research was supported by Research Grant R01 HD 04146 from the National Institute of Child Health and Human Development. The author wishes to thank Patricia Truman for assistance in testing the children and Mary O'Brien, Principal of the Martin Luther King, Jr., School, for her cooperation and aid in making *S*'s available.

[2] Requests for reprints should be sent to Peter D. Eimas, Walter S. Hunter Laboratory of Psychology, Brown University, Providence, Rhode Island 02912.

In the past, attentional explanations of the ORE were never directly tested, in that measurement of observing behavior (and presumably attention) directly and independently of the instrumental response was not obtained. This deficiency was corrected in the present study by use of an apparatus that required *S*s to press one of two observing response buttons that illuminated a single stimulus value along the chosen or attended-to dimension. The *S*s were then required to make an instrumental response to the displayed stimulus value. The choice of dimensions was that of *S*s, while the particular stimulus value that was displayed was randomly determined. In this manner, successful attainment of criterial performance demanded that *S*s learn both an observing and instrumental response. Consequently, the course of discrimination reversal learning could be analyzed in finer detail than was previously possible.

METHOD

Subjects.—The *S*s were 40 third-grade children attending public school in Providence, Rhode Island. They were randomly selected and assigned to treatments.

Apparatus and stimuli.—The apparatus consisted of two components: a stimulus and a response panel. The stimulus panel was situated in an upright position and contained four separate but adjacent display units covered with one-way mirrors. The individual display units were 1½ × 2 in. and were located approximately at eye level. In the present study, only the two left-most units were used. Directly below and attached at about a 110° angle was the response panel. Toward the top of the response panel (i.e., near the bottom of the stimulus panel) was a row of four display or observing response buttons, each of which was centered with respect to one of the display units. When pressed, a button turned on a light illuminating the stimulus in the corresponding display cell until the trial was terminated. Only the two left-most buttons were operative. Approximately 3 in. below the display buttons were two instrumental response buttons that were aligned with the two end display buttons. Two reinforcement lights were located 1 in. above the bottom response buttons and signaled the correctness of *S*'s instrumental response.

The stimuli were planometric patterns. The dimensions were color and form, with the particular values being red and green and circle and square. The stimulus values from each dimension were attached to white 5 × 8 in. cards and were spatially separated. That is, the color cue was positioned to correspond with one display window and the form cue with the second display window. The color cues were cut from Bourgess Cutocolors and were large enough to completely fill the rectangular display window. The form cues were outline ink drawings approximately 1 in. square and drawn so as to be centered in the display window. There were four stimulus cards, corresponding to the four possible combinations of stimulus values: (*a*) red in Display Unit 1 and square in Display Unit 2, (*b*) red in Display Unit 1 and circle in Display Unit 2, (*c*) green in Display Unit 1 and square in Display Unit 2, and (*d*) green in Display Unit 1 and circle in Display Unit 2.

Procedure.—Testing was done individually and was completed in a single session. Instructions were provided detailing the operation of the apparatus and the nature of the experimental task. In essence, *S* was informed that he was to learn which window to illuminate and then which of the instrumental response buttons went with each stimulus value. In addition, operation of the apparatus was demonstrated, stressing how and where the various stimulus values could appear and the consequences of correct and incorrect instrumental response choices. The *S* was also informed that he was to press only one display button per trial.

To begin a trial, *E* first inserted one of the stimulus cards into a slot on the back of the stimulus panel and signaled *S* with the word "ready" that the trial was to begin. The *S* then pressed one, and only one, display button. The choice was solely that of *S* and presumably determined by whether he wished to observe one of the colors or one of the forms. To facilitate this choice, each of the display buttons was labeled, in that directly above and below the display buttons were examples of the stimulus values that could appear in the corresponding window. Thus for the color button, a green patch was placed above the button and a red patch below the button, while for the form button, an outline drawing of a square was placed above the button and a similar drawing of a circle was placed below the button. It is important to note that while this information did not inform *S* which stimulus would appear, it did eliminate the necessity of learning which button-window combination yielded color or form cues. Upon pressing a display button, the stimulus was illuminated and remained illuminated until termination of the trial. The *S* then selected an instrumental response button, given the available stimulus information. If *S* were correct, the reinforcement light went on directly over the button pressed. If he were incorrect, the light over the other response button went on. Thus on any single trial, only one stimulus value was seen to which *S* was required to make a classification response, e.g., if red, push the left button or if green, push the right button.

Color was relevant and form irrelevant for half of the *S*s and conversely for the remaining *S*s. Within each of these groups, half of the *S*s were trained to criterion (10/10 correct observing *and* instrumental responses), while the remaining half were trained to criterion and then received 25 overtraining trials. Upon completion of the original discrimination, reversal training was begun without indication by *E* and continued until criterion was attained.

The order in which the individual stimuli along the relevant and irrelevant dimensions appeared was randomly determined, with the restriction that the same stimulus could not appear more than three times in a row. At the conclusion of the experiment, *S* was given either a package of charms or a pen.

RESULTS AND DISCUSSION

The results are shown in Table 1. A 2 × 2 analysis of variance (Dimensions × Training) of instrumental errors to criterion on the initial discrimination revealed no reliable main effects or interaction. Performance during overtraining was perfect; there was not a single observing or instrumental response error.

TABLE 1

MEAN ERRORS DURING ORIGINAL AND REVERSAL LEARNING AND MEAN IRRELEVANT OBSERVING RESPONSES DURING REVERSAL LEARNING AS A FUNCTION OF TRAINING CONDITIONS

Item	Criterion	Overtraining
Original learning		
TE	28.1	28.5
Reversal learning		
TE	7.6	2.9
PE	1.3	1.9
MRE	6.3	1.0
IOR	4.1	.5

Note.—Abbreviated: TE = total errors; PE = perseverative errors; MRE = midreversal errors; IOR = irrelevant observing responses.

An analysis of the total instrumental errors to criterion during reversal learning revealed a significant ORE, $F\ (1, 36) = 17.5, p < .001$. Neither the dimensions effect nor the Dimensions × Training interaction even approached statistical reliability. Furthermore, the manner in which overtraining operated to facilitate reversal acquisition was in agreement with predictions from attentional models of discriminative learning. First, overtraining produced a small but reliable increment in the strength of the original instrumental response as evidenced by the greater number of perseverative errors that were made by the overtrained *S*s, $F\ (1, 36) = 4.8, p < .05$. Second, overtraining significantly reduced the number of instrumental errors that occurred during the interval bounded by the end of the perseverative errors and the beginning of the criterial trials, the midreversal errors, $F\ (1, 36) = 20.2, p < .001$. One reason for the decrement in midreversal errors was that the overtrained *S*s made virtually no irrelevant observing responses. Of the 20 *S*s who received overtraining, 17 made no irrelevant observing responses, 2 made one such response, and 1 *S* made more than one irrelevant observing response. The corresponding numbers for the nonovertrained *S*s were 5, 2, and 13 *S*s. The two distributions were reliably different, $\chi^2\ (2) = 16.8, p < .001$.

Further confirmation of attentional models comes from the following analysis. After the initial perseverative errors, an *S* may do one of three things: (*a*) shift the observing response, (*b*) shift the instrumental response, but not begin the criterial trials, or (*c*) shift the instrumental response and begin the criterial run. Table 2 shows the frequency of each event as a function of the original training conditions. The effect of training was significant, χ^2 (2) = 6.9, $p < .05$, and in the direction predicted by attentional models; i.e., after overtraining there were fewer observing response shifts and more instrumental response shifts that began the criterial run.

Two final comments are in order: One concerns the competing explanations of the ORE; the second concerns the relation between the overt observing response, as measured in the present study, and the theoretical construct of attention as a central selective process (Sutherland, 1964; Zeaman & House, 1963). With regard to the former, all three competing explanations mentioned previously are unable to account for two aspects of the data. First, there is the increment in perseverative errors after overtraining. Certainly this is an unexpected event if overtraining functions to make the changed reinforcement contingencies more discriminable, or if overtraining produces a decrease in avoidance tendencies to the negative stimulus, or if overtraining results in some general response of discrimination. These explanations require that an overtrained *S* shift his instrumental response more quickly after encountering nonreward. Second, these explanations are unable to deal with the observing response data. Attentional models, on the other hand, are explicit in predicting these effects in addition to the others reported.

TABLE 2

Frequency with Which *S*s after the Initial Perseverative Errors Shifted the Observing Response (SOR), Shifted the Instrumental Response, but Did Not Begin the Criterial Run (SIR), or Shifted the Instrumental Response and Began the Criterial Run (SIRC) as a Function of Original Training

Item	SOR	SIR	SIRC
Criterion[a]	8	5	4
Overtraining	3	4	13

[a] There were three *S*s in this group who could not be categorized because they made no perseverative errors, in that the initial response was correct although errors did occur before criterion was obtained.

With respect to the relation between observing responses as defined by the present experimental situation and the theoretical construct of attention, it has been assumed that the former was a behavioral representation of the latter. The fact that the observing response behavior mirrored virtually perfectly the expected functioning of attention lends strong support to this assumption. As a consequence, an easily measurable correlate of attention is available that may provide a means for further examination of the role of this central selective process in a variety of problem-solving tasks.

REFERENCES

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts of training. *Journal of Comparative and Physiological Psychology,* 1957, **50**, 195–198.

D'Amato, M. R., & Jagoda, H. Analysis of the role of overlearning in discrimination reversal. *Journal of Experimental Psychology,* 1961, **61**, 45–50.

Eimas, P. D. Attentional processes. In H. W. Reese & L. P. Lipsitt (Eds.), *Experimental child psychology.* New York: Academic Press, in press.

Lovejoy, E. Analysis of the overlearning reversal effect. *Psychological Review,* 1966, **73**, 87–103.

Mackintosh, N. J. Selective attention in animal discrimination learning. *Psychological Bulletin,* 1965, **64**, 124–150.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *Journal of Experimental Psychology,* 1953, **46**, 107–112.

Sutherland, N. S. The learning of discrimination by animals. *Endeavour,* 1964, **23**, 148–152.

Wolff, J. L. Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin,* 1967, **68**, 369–408.

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963.

(Received April 14, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 503–510

TRANSFER OF EYELID CONDITIONING FROM INSTRUMENTAL TO CLASSICAL REINFORCEMENT AND VICE VERSA[1]

DAVID A. GRANT,[2] NEAL E. A. KROLL,[3] BARRY KANTOWITZ,
MICHAEL J. ZAJANO,[4] AND KENNETH B. SOLBERG[5]

University of Wisconsin

Two parallel experiments investigated transfer to instrumental eyelid conditioning following 0, 5, 10, or 20 classical reinforcements with the same CS and transfer to classical eyelid conditioning following 0, 5, 10, or 20 instrumental reinforcements that were contingent on *S*s emitting an eyelid response that resembled a classically conditioned CR upon the appearance of the CS. Amount of transfer to instrumental conditioning was a progressively increasing function of the number of pretransfer classical reinforcements, and the instrumental acquisition curves were similar to those previously obtained. Also, the conditioned (C) or voluntary (V) classification of the instrumental response (IR) form resembled the response form that was recorded during pretransfer training. Transfer from instrumental to classical conditioning was not influenced by the number of pretransfer instrumental reinforcements, the classical acquisition curves were abnormal, and the response form in the transfer phase was not related to the response form recorded in the pretransfer phase. It is concluded that transfer from classical to instrumental training occurs readily and is a function of the number of pretraining classical reinforcements, but that a parallel transfer from instrumental to classical conditioning may be more difficult to demonstrate if it does occur.

The two experiments reported here investigated transfer to instrumental conditioning following varied numbers of classical reinforcements and transfer to classical conditioning following varied numbers of instrumental reinforcement.

The relationship between classical and instrumental conditioning is a matter of continuing interest (see, e.g., Kimble, 1961). Various strategies that have been employed in attempting to establish the relations between the consequences of these two basic reinforcing operations are concisely described by Tighe, Graves, and Riley (1968). Longer reviews outlining the complexity of the investigative program are given by Kimmel (1967) and Rescorla and Solomon (1967). To summarize experimental results briefly, it would appear that both smooth muscle and skeletal muscle responses can be conditioned by either classical or instrumental reinforcement, but the possibility of the mediation of the recorded response by other responses is never really resolved. Investigations of the similarity of functions yielded by the two kinds of learning (e.g., delay of reinforcement) show many parallels between outcomes of classical and instrumental reinforcement experiments, but there are some exceptions to these parallels (see Kimble, 1961, Ch. 4).

Transfer of conditioning of responses of the same effector system from instrumental to classical reinforcement or from classical to instrumental reinforcement would seem to be an attractive approach to add information on the more general problem of the relationship between the two learning processes. The extent to which the availability and stimulus control of a response that have been brought about by classical reinforcement will contribute to further conditioning to the same CS when the mode of reinforcement is changed from classical to instrumental and vice versa would appear to depend relatively directly on the relationship between the two learning processes and the identity of the responses involved. Presumably, the more closely the processes and responses are functionally related, the greater should be the transfer.

---

[1] This research was supported in part by the United States Public Health Service through Grant MH06792 to the first author. The first experiment was planned by the first, second, and fourth authors and executed by the latter two; the second experiment was planned by the first and third authors and executed by the third and fifth. The first author takes primary responsibility for the form of this report.

[2] Requests for reprints should be sent to David A. Grant, Department of Psychology, Psychology Building, Charter at Johnson, University of Wisconsin, Madison, Wisconsin 53705.

[3] Former United States Public Health Service predoctoral fellow, now at the University of California.

[4] National Science Foundation Undergraduate Research Participation student, Summer 1967.

[5] Now at St. Mary's College, Winona, Minnesota.

Positive transfer from classically conditioned responses to instrumentally conditioned responses has been reported frequently (e.g., DeBold, Miller, & Jensen, 1965; Ellison & Konorski, 1965; Konorski, 1967). Similarly, transfer of differential classical conditioning to differential instrumental conditioning has been reported by Bower and Grusec (1964), Trapold (1966), and Trapold, Lawton, Dick, and Gross, 1968). Transfer from instrumental reinforcement to classical reinforcement seems, however, rarely or never to have been employed. Although a complete parallel between classical training and instrumental training procedures is impossible to achieve, the present two experiments were matched closely so as to give at least preliminary evidence on the question of the degree to which transfer of instrumental to classical training parallels that of transfer of classical to instrumental training.

The eyelid response is especially well suited to this investigation for two reasons. First, the eyelid can be conditioned by classical, instrumental avoidance, instrumental reward, and combinations of these methods (see, e.g., Hansche & Grant, 1965). Second, the latency and topography of the eyelid response is well displayed on the recordings, and although some debate on the matter persists (Goodrich, 1966), there is little question but that the voluntary (V) and conditioned (C) form of the eyelid CR (Spence & Taylor, 1951) can be objectively differentiated (Hartman & Ross, 1961). With the general experimental parameters used in our laboratory, instrumentally trained *S*s give predominantly the V form of the eyelid CR, and classically trained *S*s split approximately evenly between V- and C-form responders if the Hartman-Ross response slope criterion is employed. It would, therefore, appear to be feasible to discover whether the transfer from one mode of reinforcement to another is such that the response form emitted by or elicited from *S* in the pretransfer phase is the same as the response form emitted or elicited from *S* in the posttransfer phase. Although not completely definitive, this comparison should provide some evidence as to whether the response conditioned and made available in the pretransfer phase is the same response that continues to be emitted and conditioned during the posttransfer phase.

Aside from evidence on the specific transfer between modes of reinforcement, the results of such experimentation should be more generally relevant to investigators interested in the Soviet psychology of volition, which depends in part on the notion that a classically conditioned response can readily come under control of verbal stimuli and thus be modified, more highly differentiated, and more subject to stimulus generalization through the second signaling or language system that would involve all of the interrelations available in the language repertoire of *S* (see, e.g., Luria, 1961).

EXPERIMENT I

TRANSFER FROM CLASSICAL TO INSTRUMENTAL TRAINING

Method

Apparatus.—The apparatus and laboratory were the same as that used by Fleming and Grant (1966) and Cerekwicki and Grant (1967). The *S* sat in an ophthalmological examination chair in an inner soundproofed chamber and faced a rectangular wooden enclosure painted flat white and illuminated to about 2 mL. At the back of the enclosure, approximately 120 cm. in front of *S*'s eyes, were two milk-glass disks, 10 cm. in diameter, whose centers were separated by 66 cm. Between the milk-glass disks was a 45 × 30 cm. ground-glass screen that was not used in the experiment. The prestimulation illumination of the disks was about .01 mL. The *S* wore a headset carrying a microtorque potentiometer to pick up the eyelid movement and a jet from which the corneal puff could be delivered.

The CS was always the simultaneous illumination of the two ground-glass disks to .08 mL. for a duration of 1 sec. The classical UCS, when given, was a 200-msec. corneal puff delivered 660 msec. after CS onset. The puff intensity was regulated at 3 psi. The instrumental reinforcement, when given, was the same as that used by Cerekwicki and Grant (1967). It consisted of a 660-msec. illumination of a green 7½-w. lamp and the operation of a Veeder-Root counter. The instrumental reinforcement was given 2 sec. after an instrumental response (IR), which was defined as a 1-mm. or greater eyelid closure occurring 200–660 msec. after CS onset. Detection of the IR was arranged to be automatic by means of a Schmitt trigger and associated timing circuitry.

For each *S*, the experiment was divided into two phases, a classical conditioning phase and an instrumental conditioning phase. The *S*s were instructed to keep looking at the ground-glass disks at all times. Before the classical conditioning phase of the experiment, *S*s were read standard neutral instructions, which asked them not to aid or inhibit their natural eyelid responses. During this phase of the experiment, *S*s were divided into four groups who received, respectively, 0, 5, 10, or 20 classically reinforced pairings of the CS and the UCS. The intertrial interval varied from 15 to 35 sec., averaging 25 sec. Upon completion of the first phase of the experiment, all *S*s were instructed regarding the instrumental reinforcement in the same fashion as those *S*s employed in the experiments of Hansche and Grant (1965), Fleming and Grant (1966), and Cerekwicki and Grant (1967). They were told that their object now was to make the green light come on and the counter to count on each trial, but they could not be told "what you have to do to make the green light come on." Then followed 60 trials during which all IRs were reinforced automatically.

Subjects.—Thirty-two men and 48 women students served as *S*s in the experiment. Assignment of 20 *S*s to each of the four experimental conditions was random except that the sex ratio was balanced.

Results

An IR was defined as any eyelid response of 1 mm. or greater magnitude occurring between 200 and 660 msec. after the onset of the CS. The IRs were automatically recorded by the electronic pick-off unit and were later hand-checked from this Oscillo/riter record.

General transfer effects.—Overall transfer from classical to instrumental conditioning is shown in Fig. 1, where percentage of IRs over all trials, and total mean amplitude of IRs are plotted as functions of the number of pretransfer classical reinforcements. It is clear from the curves in Fig. 1 that classical reinforcement transfers to the subsequent instrumental training, because with increasing numbers of classical reinforcements, both the number and amplitude of the IRs during the posttransfer phase increase. The data were analyzed by orthogonal polynomials (Grant, 1956) using the arc-sine transformation for percentages. This analysis showed that the linear component of the trend was highly significant, F (1, 76) = 8.49, $p < .005$. The higher order components were not significant, indicating that the trend was primarily linear. A similar analysis of the amplitude data gave similar results, the linear component being significant, F (1, 76) = 5.69, $p < .05$. Frequency and amplitude were significantly correlated, r (78) = .40, $p < .001$.

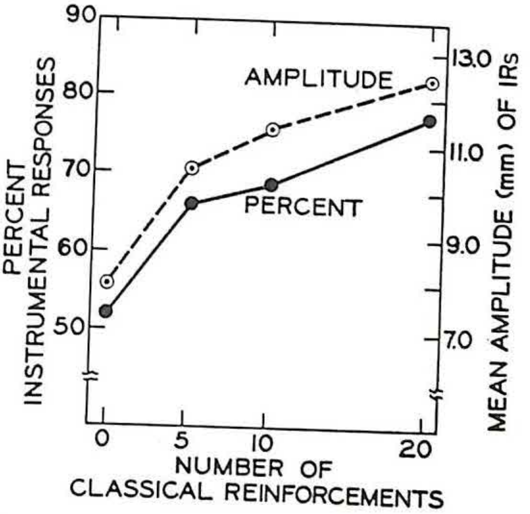


FIG. 1. Percentage frequency of instrumental eyelid responses and their mean amplitudes during the transfer phase of Exp. I as a function of the number of pretransfer classically reinforced training trials.

The acquisition curves for the IRs are shown in Fig. 2. Here percentage of IRs is plotted for each successive 10-trial block, with the number of classical reinforcements in the pretransfer phase being the parameter. The curves are approximately parallel negative exponential growth functions with no initial inflection. The absence of an initial inflection for the zero-classical-reinforcement curve is especially curious. This is

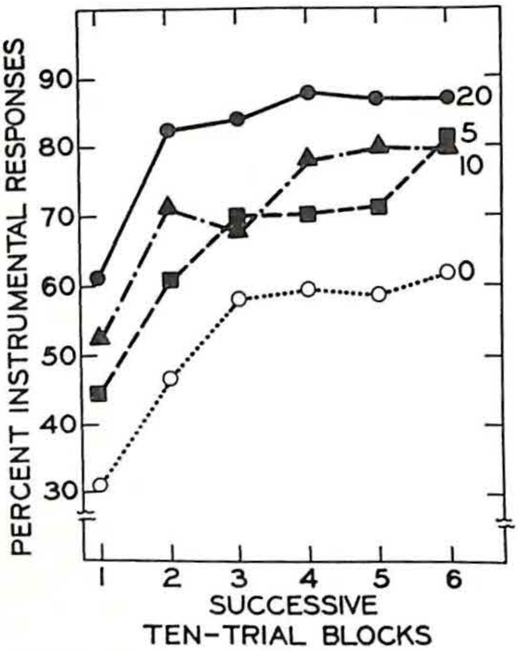


FIG. 2. Acquisition of the instrumentally reinforced eyelid response during the six successive 10-trial blocks of the transfer phase of Exp. I. (The parameter is the number of pretransfer classically reinforced training trials.)

true because an instrumental reinforcement and consequent increment in the acquisition of the instrumental response will supposedly occur only if a response is emitted. Therefore, when the initial emission rate of the response is low (e.g., 30% on the first 10-trial block of the group having received zero classical training trials), instrumental reinforcement would occur sparingly and the initial rate of learning the instrumental response would be slow, whereas when the response emission rate is high (e.g., 60% on the first 10-trial block of the group having received 20 classical reinforcements or the second 10-trial block of the group having received 5 classical reinforcements), the learning rate should be twice as rapid. In a group such as that having received zero classical reinforcements, there should then be an initial period of slow learning followed by more rapid learning as the response emission rate increases, i.e., there should be an inflection in the learning curve. The curious point is that this did not occur. If the data are plotted in terms of the number of IRs per 10-trial blocks divided by the number of IRs for the previous 10-trial block, it is clear that the early reinforcements have a greater effect on the increase in response rate than do the reinforcements later in the training.

Analysis of the latency data indicated that the number of classical reinforcements in the pretransfer phase had little effect on the latency of the subsequent IRs. The curve was essentially flat.

Response form in transfer.—In Table 1, the number of *S*s whose response forms could be classified as C or V is given for the 5-, 10-, and 20-classical-reinforcement groups. Then in the body of the table, the average number of C-form and V-form IRs are given for the Cs and Vs of these groups. It is clear from Table 1 that the V form predominated during instrumental conditioning. There is also a clear tendency for *S*s giving the C form during classical training to give relatively more C-form IRs than *S*s showing the V form of the eyelid CRs during classical training. Furthermore, the more classical training trials that were given, the stronger this tendency for C-form responders to give C-form IRs. These trends in the data were borne out by the multiple chi-square analysis (Winer, 1962, pp. 631–632), but the chi-square analysis can only be suggestive because of violations of assumptions. The results do, however, indicate that *S*s show some consistency in giving the same response form in both the classical and the instrumental phases of the experiment.

TABLE 1

MEAN NUMBER OF V- AND C-FORM IRS DURING THE INSTRUMENTAL CONDITIONING PHASE FOR *S*S CLASSIFIED AS CS AND VS DURING THE PRETRANSFER CLASSICAL TRAINING PHASE OF EXP. I

Response forms in classical phase	*N*	C-form IRs	V-form IRs
Group 5			
C	6	8.33	34.33
V	4	14.75	30.25
Group 10			
C	12	15.67	20.83
V	6	5.00	49.17
Group 20			
C	8	23.62	21.38
V	8	3.33	50.25

Experiment II

Transfer from Instrumental to Classical Training

Method

Apparatus.—The laboratory, apparatus, and stimulation procedures were exactly the same as in Exp. I, except that the light for the instrumental reinforcement was white.

Procedure.—In this experiment, an instrumental training phase preceded the classical training phase. After receiving the instrumental reinforcement instructions, three groups of *S*s were trained until they received either 5, 10, or 20 instrumental reinforcements, respectively, contingent on their emitting IRs to the CS. A maximum of 60 training trials was given, and 13 *S*s who did not reach the criterion number of IRs in 60 trials were discarded. The fourth group of *S*s received either 5, 10, or 20 noncontingent instrumental reinforcements on a schedule corresponding to that received by 6 *S*s randomly selected from each of the three instrumental reinforcement groups. Thus the zero group did receive instrumental reinforcement, but it was not contingent on performing IRs. After the instrumental phase of the experiment, all *S*s were given 60 classical training trials, combining the CS with the UCS stimulation but with no change in instructions. At this point in the experiment, the instrumental reinforcing white light and counter functions were terminated. Finally, all *S*s received 30 extinction trials with the CS alone.

Subjects.—The four groups of 18 *S*s each consisted of 24 men and 48 women students from introductory classes. Assignment was random until it became necessary to balance for sex. In addition to *S*s discarded for failure to obtain the specified number of instrumental reinforcements within 60 trials, 6 *S*s were discarded because of assorted mechanical difficulties, and one was discarded because he failed to give any UCRs during classical training.

Results

Response form and transfer.—Records were scored according to the same criteria used in Exp. I. During instrumental training, the mean numbers of nonreinforced instrumental trials were 11.5, 12.2, and 10.6 for the 5-, 10-, and 20-IR groups, respectively, indicative of the rapidity of learning. All *S*s in the 5-, 10-, and 20-IR groups could be classified as V- or C-form responders during instrumental training and subsequently reclassified during classical training according to the Hartman and Ross (1961) criterion. Table 2 shows the relationship between the V-C classification during instrumental training and the V-C classification during classical training. Forty-five *S*s were classified as V-form responders during instrumental training, and 36 *S*s were classified as V-form responders during classical training. Thirty-seven *S*s had the same response form classification during both instrumental and classical training, while 17 *S*s shifted form. An insignificant χ^2 (1) = 1.35 indicated that there was no significant tendency for the response form during instrumental training to correspond to the response form during classical training, as the cell frequencies were quite accurately predicted by the marginals.

TABLE 2

Response Form Classification of *S*s during Pretransfer Instrumental Training and during Subsequent Classical Training in Exp. II

Classification during instrumental training	Classification during classical training		Totals
	C	V	
V	13 (15)	32 (30)	45
C	5 (3)	4 (6)	9
Totals	18	36	54

Note.—Expected frequencies are in parentheses.

General transfer effects.—In Fig. 3 are plotted the percentages of CRs during classical training as a function of the number of instrumental reinforcements in the pretransfer phase. These are plotted separately for the Vs and Cs during the classical acquisition and extinction trials. It is apparent that in constrast to transfer from classical to instrumental training, there is here little evidence of progressively greater transfer from instrumental to classical conditioning as a function of the number of pretransfer instrumental reinforcements. Analysis of variance on both acquisition and extinction measures gave insignificant *F*s throughout for all independent variables, the largest *F* (1, 64) = 1.62.

The acquisition curves for the different groups during classical training are plotted

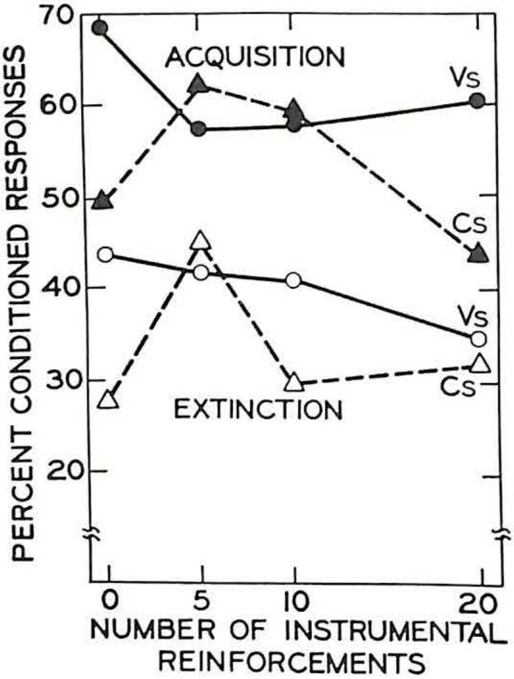


FIG. 3. Percentage frequency of eyelid CRs during the acquisition and extinction trials of the transfer phase of Exp. II. (Data of C-form and V-form responders are plotted separately.)

in Fig. 4. Percentages of CRs are plotted for successive 10-trial blocks in the upper panel for the Vs and in the lower panel for the Cs. The parameter is the number of contingent instrumental reinforcements in the pretransfer phase. It is clear that the Vs start with a high initial rate of responding and show very little tendency to increase the percentage frequency of CRs during classical training. The Cs also start with a relatively high rate of responding and show some acqusition trends. The relatively high initial rates are probably accounted for by the fact that all *S*s had initially been given the instrumental training instructions, and there were no changes in the instructions when the experiment shifted from the instrumental training phase to the classical training phase. It is especially noteworthy that the zero-instrumental-reinforcement group, which had received no contingent instrumental reinforcement, showed as good or better classical acquisition as those groups that had received instrumental pretraining. All of this indicates that the number of contingent instrumental reinforcements had little effect on subsequent classical conditioning.

When the percentages of CRs on the first 10 trials of classical acquisition were plotted trial by trial, it appeared that the very high level of responding for both the Cs and the Vs who had 10 or 20 instrumental reinforcements was quickly disrupted with the onset of classical training. With the slightly aversive classical UCS, the curves for the 10- and 20-IR groups showed an initial drop after the first trial, which in the case of the C-form responders in the 20-IR groups went

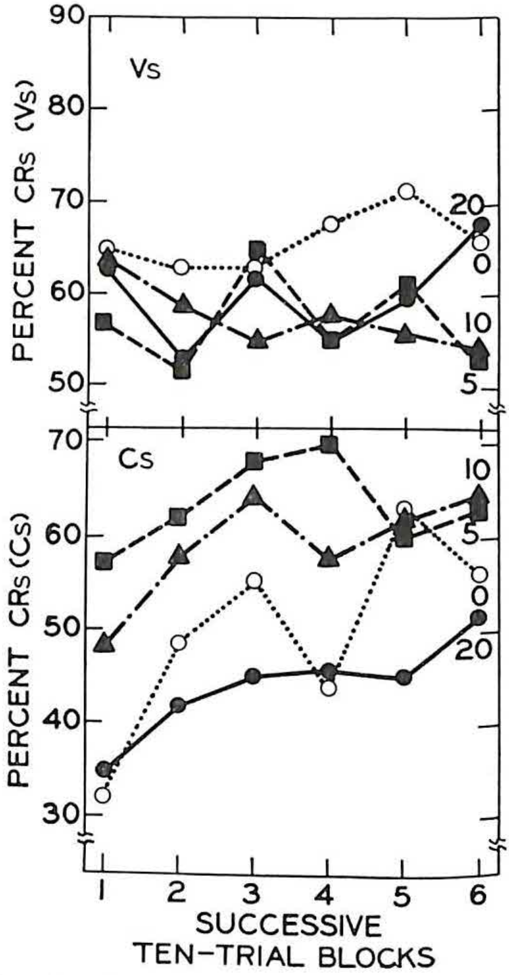


FIG. 4. Acquisition curves during the six successive 10-trial blocks of classical reinforcement during the transfer phase of Exp. II. (Data of the V-form responders are given in the upper panel; those of the C-form responders in the lower panel. The parameter is the number of instrumental reinforcements contingent on performance of an eyelid response during the pretransfer phase of the experiment.)

all the way to zero CRs by the sixth trial. Thus it would appear that the transition from instrumental to classical training is a far different process than the transition from classical to instrumental training.

Discussion

The results of the two experiments appear to show an asymmetry in transfer between the two forms of reinforcement. The amount of subsequent instrumental conditioning became progressively greater with greater numbers of classical reinforcements in the pretransfer phase. The acquisition functions during instrumental training were similar to those found previously and were ordered by the number of pretransfer classical reinforcements. There was also a clear tendency for the topography of the IRs in the transfer phase to be the same as the topography of the classical CRs that appeared in the pretransfer phase when 10–20 classical reinforcements had been given. In contrast, neither the C- nor the V-form responders during prior instrumental training showed any tendency for the amount of subsequent classical conditioning to be a function of the number of pretransfer instrumental reinforcements. Only the C-form responders during the classical training phase showed anything like the normally obtained increments in response rate or learning curves during the instrumental training phase (compare the upper and lower panels of Fig. 4). There was considerable initial transfer for both Cs and Vs as they made the transition from the instrumental to the classical training phase. This, however, could hardly be attributed to the instrumental training per se, but more likely was a function of the instrumental instructions because even the yoked noncontingent control *S*s started instrumental training with a high response rate. Also, there was no evidence that *S*s giving the V or C forms of response in the prior instrumental conditioning tended to give the same response form in the subsequent classical conditioning.

The implications of these findings appear to be that classical reinforcement, however it may operate, makes some sort of eyelid response available for subsequent instrumental reinforcement and will enhance the effect of that subsequent training. It is quite likely that during classical training the conditioning process may differ from *S* to *S*. Aside from the classical conditioning that should occur in eyelid conditioning, where a CS is followed by the corneal air puff, an anticipatory CR partially avoids the aversive character of the UCS, so that some instrumental avoidance training may take place. Therefore, it would appear that the eyelid CR may be conditioned: (*a*) exclusively by classical reinforcement; (*b*) initially by classical reinforcement and subsequently by avoidance instrumental reinforcement; or (*c*) initially emitted as an operant, it is then conditioned by avoidance training. Conceivably, the same *S* might become conditioned by a complex mixture of these processes, but the end result is that he has an eyelid response available to perform during subsequent instrumental training so that the instrumental reinforcement becomes quickly effective in its own right. When, however, the transition is made from initial instrumental training to classical training, perhaps only those *S*s who are, in effect, learning by instrumental avoidance during the subsequent classical training may benefit from the number of instrumental appetitive reinforcements in the pretransfer phase.

Nevertheless, the failure of progressively increased number of instrumental reinforcements to progressively facilitate subsequent classical conditioning remains a puzzle. Other experiments on this problem appear not to have been done, and it is possible that other types of transition from instrumental to classical training or the employment of other responses might give different results. For example, in Exp. II there was no change in the instrumental training instructions before the classical training began. This meant, among other things, that the standard neutral classical training instructions were not given, so that *S* may have had a relatively more active attitude during the classical training than he would have had were he given neutral instructions at the transition. This could, of course, account for the relatively high initial levels during classical training found with all Vs and several groups of Cs.

Three other features of the transition from classical to instrumental training may have been important. First, with the termination of instrumental reinforcement, the white light and counter no longer operated, so that there may have been some extinction of the IR that would interfere with the classical conditioning. Second, the advent of the new aversive stimulus, the UCS, may have had a disturbing effect on *S*'s behavior. Either of these processes should produce some initial reduction in eyelid responding during the first 10 classical training trials, and this was indeed observed for *S*s who had had 10 or 20 instrumental reinforcements in the pretransfer phase. Any transfer from one reinforcing procedure to another is likely to be somewhat disruptive, and the particular form of the transition may greatly influence the course of subsequent transfer performance. The third, and possibly most important, source of asymmetry in transfer has to do with a difference between classical and instrumental CR topography in relation to avoiding the aversiveness of the air puff. In transfering from classical to instrumental reinforcement, the topography of the pretransfer classically conditioned response was of little import, because with the Schmitt trigger setting in the instrumental phase, both large CRs and small CRs would receive instrumental reinforcement if they were emitted during the appropriate latency interval. With initial instrumental training, however, many *S*s learned a minimal, highly stereotyped CR that was quite ineffective in avoiding the puff during the subsequent classical training phase. Following 10–20 instrumental reinforcements, such an abortive response may be difficult to alter or extinguish and may interfere with the classical conditioning process. Thus the different response topographies would favor transfer from classical to instrumental training and perhaps interfere with transfer from instrumental to classical training.

The above considerations make clear that to draw a general conclusion that the amount of pretransfer instrumental training does not affect subsequent classical conditioning would be premature.

REFERENCES

BOWER, G., & GRUSEC, T. Effect of prior Pavlovian discrimination training upon learning and operant discrimination. *Journal of the Experimental Analysis of Behavior,* 1964, **7**, 401–404.

CEREKWICKI, L. E., & GRANT, D. A. Delay of positive reinforcement in instrumental eyelid conditioning. *Journal of Experimental Psychology,* 1967, **75**, 360–364.

DEBOLD, R. C., MILLER, N. E., & JENSEN, B. D. Effect of strength of drive determined by a new technique for appetitive classical conditioning. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 102–108.

ELLISON, G. D., & KONORSKI, J. An investigation of the relations between salivary and motor responses during instrumental performance. *Acta Biologiae Experimentalis,* 1965, **25**, 297–315.

FLEMING, R. A., & GRANT, D. A. A comparison of rate and contingency of classical and instrumental reinforcement upon the acquisition and extinction of the human eyelid CR. *Journal of Experimental Psychology,* 1966, **72**, 488–491.

GOODRICH, K. P. Elemental analysis of response slope and latency as criteria for characterizing voluntary and nonvoluntary responses in eyeblink conditioning. *Psychological Monographs,* 1966, **80**(14, Whole No. 622).

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychological Bulletin,* 1956, **53**, 141–154.

HANSCHE, W. J., & GRANT, D. A. A comparison of instrumental reward and avoidance training with classical reinforcement technique in conditioning the eyelid response. *Psychonomic Science,* 1965, **2**, 305–306.

HARTMAN, T. S., & ROSS, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *Journal of Experimental Psychology,* 1961, **61**, 334–338.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

KIMMEL, H. D. Instrumental conditioning of automatically mediated behavior. *Psychological Bulletin,* 1967, **67**, 337–345.

KONORSKI, J. *Integrative activity of the brain: An interdisciplinary approach.* Chicago: University of Chicago Press, 1967.

LURIA, A. R. *The role of speech in the regulation of normal and abnormal behavior.* New York: Liveright, 1961.

RESCORLA, R. A., & SOLOMON, R. L. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychological Review,* 1967, **74**, 151–182.

SPENCE, K. W., & TAYLOR, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *Journal of Experimental Psychology,* 1951, **42**, 183–188.

TIGHE, T. J., GRAVES, E. M., & RILEY, C. A. Successive reversals of a classically conditioned heart-rate discrimination. *Journal of the Experimental Analysis of Behavior,* 1968, **11**, 199–206.

TRAPOLD, M. A. Reversal of operant discrimination by non-contingent discrimination reversal training. *Psychonomic Science,* 1966, **4**, 247–248.

TRAPOLD, M. A., LAWTON, G. W., DICK, R. W., & GROSS, D. M. Transfer of training from differential classical to differential instrumental conditioning. *Journal of Experimental Psychology,* 1968, **76**, 568–573.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received April 15, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 511–514

PROACTIVE INHIBITION AND UNDETECTED RETENTION INTERVAL REHEARSAL[1]

JOHN P. HOUSTON[2]

University of California, Los Angeles

The *S*s learned either a single A-C list or successive A-B, A-C lists. Half of the *S*s in each of these conditions were told, following learning, that the experiment was finished. The remaining halves were asked to return in 1 wk. All *S*s were then contacted by mail and asked to recall the second-list materials. Significant proactive inhibition (PI) appeared when *S*s were expecting to return to the experimental setting. No PI appeared when *S*s believed the experiment to be finished. The *S*s reporting greater degrees of retention interval rehearsal (RIR) displayed greater amounts of PI. It was suggested that PI is determined not by competition at the time of recall, but by unlearning associated with RIR.

Significant proactive inhibition (PI) has been found with the modified-modified free recall (MMFR) test (Houston, 1967; Koppenaal, 1963; Wipf & Webb, 1962). According to the two-factor theory of forgetting, this should not be the case. PI is presumed to be determined by competition at the time of recall and the MMFR test has been assumed to eliminate these kinds of competition effects. Hence, no PI should appear when this recall test is used.

At least two explanations of this disparity have appeared. First, the MMFR task may not be completely effective in eliminating competition effects (Postman, Stark, & Fraser, 1968). If this proves to be the case, then previous distinctions between unlearning and competition (Barnes & Underwood, 1959) become tenuous. Second, something in addition to competition at the time of recall may contribute to obtained PI effects. Keppel (1968) suggested that the undetected unlearning of second-list associations during the retention interval may be one such factor. Houston (1968) proposed that this kind of unlearning may be associated with uncontrolled rehearsal occurring during the retention interval. The undetected rehearsal of the materials during the retention interval could account for the loss of some of the second-list items and the maintenance of the first-list items over time. That is, first-list items which were strong enough to resist unlearning during the formal learning of the second list should be strong enough to resist whatever informal unlearning might tend to occur during the retention interval. On the other hand, the second-list items have not undergone any unlearning at the beginning of the retention interval. Thus they may be more susceptible to whatever unlearning events occur during the interval.

The present experiment was concerned with the latter interpretation which implies that PI is determined by rehearsal and unlearning and not by competition. If PI is due to unlearning associated with undetected retention interval rehearsal (RIR), then a procedure which minimizes rehearsal during the retention interval might also reduce PI effects. Accordingly, one group of *S*s in the experiment volunteered in the normal fashion. That is, they believed that 1 wk. after the original learning session they were to return to the experimental setting for a second session. A second group did not know they were going to be tested for recall after 1 wk. They were dismissed following the learning session and told that the experiment was finished. All *S*s were then contacted by mail after 1 wk. They were asked to complete and return by mail an enclosed recall test sheet. The general prediction was that the groups which believed they were to return after 1 wk. would rehearse more dur-

---

[1] This work was supported by United States Public Health Service Grant MH 15336-01.

[2] Requests for reprints should be sent to John P. Houston, Department of Psychology, University of California, Los Angeles, California 90024.

ing the week and would display greater PI effects on the recall test than groups which believed the experiment was finished.

METHOD

Subjects.—The experiment involved a 2×2 design, with half of the *S*s learning successive A-B, A-C lists and the other half learning a single A-C list. Half of each of these groups was asked to return, while the remaining halves were dismissed. Each of the four conditions contained 24 *S*s drawn from the psychology department *S* pool. The four conditions were unsystematically arranged such that each occurred once in 24 successive blocks of *S*s. The *S*s were assigned to this arrangement as they appeared in the laboratory. During the course of the experiment, 32 *S*s were replaced. Of these replacements, 22 were due to the fact that correct addresses could not be obtained following the learning session. Addresses were taken from university listings, which were in some cases either incomplete or out of date. The remaining replaced *S*s failed to return their recall sheets. The numbers of replacements in the experimental no-return, experimental return, control no-return, and control return conditions were 7, 8, 8, and 9, respectively.

Materials.—The materials in the present study were the A-B and A-C lists developed by Barnes and Underwood (1959). A detailed description of these CVC-adjective lists is contained in their article. One list always served as the A-B list, while the other always served as the A-C list. Two pairings of the stimulus and response materials were developed and used equally frequently in each condition. Three unsystematic presentation orders were arranged and each was used equally frequently as the starting order.

Procedure.—All learning was conducted at a 2:2-sec. presentation rate, with 4-sec. intertrial and 45-sec. interlist intervals. A Stowe memory drum was employed. The *S*s in the experimental, or two-list, conditions first learned an A-B list to a criterion of 1 perfect trial. All *S*s then learned the A-C list for 15 anticipation trials. Following learning, the return *S*s were dismissed, with the understanding that they were to return in 1 wk. The no-return *S*s were led to believe that the experiment was finished and were dismissed. Five days after the learning session, all *S*s were mailed recall forms. The stimulus materials were listed in a new random order on these forms. The *S*s were asked to recall the responses and record them in the appropriate spaces. In addition, a questionnaire concerning *S*'s activities during the week was included. The *S*s were asked to return their completed forms in an enclosed, self-addressed envelope.

RESULTS

The mean numbers of trials required by the experimental return (R) and experimental no-return (NR) groups to reach criterion on the first list were 9.20 and 9.08, respectively, $t < 1$. The four conditions did not differ in terms of mean numbers of correct second-list learning responses, *F*s $(1, 92) < 3.83$. The mean retention intervals, based on the times *S*s recorded in connection with their recall tests, were 7.37, 7.90, 7.83, and 7.54 days for the experimental NR, experimental R, control NR, and control R groups, respectively. These means did not differ significantly, *F*s $(1, 92) < 2.57$. Each *S* was asked to record the time he began and the time he finished the recall test. The mean numbers of minutes used by the experimental NR, experimental R, control NR, and control R groups were 13.29, 15.26, 12.37, and 10.58, respectively. These means did not differ significantly, *F*s < 2.05. All *S*s reported finishing the recall test in one session. These results suggest that the mailing technique was quite successful in producing uniform recall tests.

The mean numbers of second-list items recalled correctly were 4.50, 4.21, 3.38, and 5.33 for the experimental NR, control NR, experimental R, and control R conditions, respectively. The NR-R main effect was not significant, $F < 1$. Recall in the control conditions was significantly greater than recall in the experimental conditions (one-tailed test), $F\ (1, 92) = 3.28, p < .05$. More important, the interaction was significant, $F\ (1, 92) = 5.98, p < .05$. These results suggest that significant PI appeared in the R conditions, but that none appeared in the NR situation. Subsequent *t* tests revealed that the control and experimental R groups did differ, $t\ (46) = 7.50, p < .01$, while the control and experimental NR groups did not differ, $t < 1$.

The mean numbers of A-B responses recalled correctly by the experimental R and experimental NR groups were 3.00 and 2.24, respectively. Although not significant ($t = 1.46$), this difference is in line with the notion that more rehearsal of the A-B associations should occur in the return condition.

The lack of a significant difference could be related to the suggestion that the A-B associations which remain following A-C learning are quite strong. It may have been because of this kind of ceiling effect that informal practice of these associations during the retention interval did not result in an increase in recall performance.

The questionnaire data indicated that 43 of the 48 *S*s in the R conditions felt, prior to the receipt of the questionnaire, that they would be asked to recall the materials upon returning to the laboratory setting. Only 5 of the NR *S*s indicated that they felt they would be asked to recall the materials at some time. Seven R *S*s and 3 NR *S*s admitted writing some of the materials down during the retention interval.

One of the questions asked how many of the items *S* had thought of or rehearsed during the interval. Another item asked how often *S* had thought of these items. The product of these two estimates provides a very rough index of the number of item rehearsals accomplished by each *S*. The mean numbers of estimated rehearsals for the experimental NR, control NR, experimental R, and control R groups were 6.37, 6.20, 9.83, and 10.38, respectively. Although these means did not differ significantly, $F = 2.36$, the R conditions did produce more rehearsals than the NR situation.

Another way to look at the relationship between interference effects and retention interval rehearsal is to split *S*s into high- and low-rehearsal groups and then compare the amounts of PI obtained in these categories. Accordingly, *S*s with rehearsal scores of 10 and above were included in a high-rehearsal category, while *S*s with unit rehearsal scores of 0 constituted a low-rehearsal category. The mean second-list recall scores of the experimental and control high-rehearsal *S*s were 4.15 and 5.70, respectively, $t\ (21) = 2.09$, $p < .05$. The mean recall values in the corresponding low-rehearsal groups were 3.90 and 4.20, $t < 1$. These means are in agreement with the assumption that retention interval rehearsal is associated with greater interference effects. The loss in the high-rehearsal conditions was 1.55 items, while the loss in the low-rehearsal situation was .3 items.

DISCUSSION

The principal findings of the present study were that significant PI appeared in the R situation, but that none occurred in the NR condition. PI seems to be related to *S*'s perception of future events. Significant PI occurred when *S*s felt that there was a good possibility that they would be tested for recall. When *S*s were convinced that the experiment was over and that they would not be tested for recall, no PI appeared.

There are, of course, a number of ways to explain these findings. For example, if one wished to maintain the traditional interpretation of PI based on the notions of spontaneous recovery and competition at the time of recall, one might argue as follows. First, spontaneous recovery may be related to the strength of *S*'s suspicion concerning the likelihood of an ensuing recall test. If this were true, more recovery would be expected in the R condition. Second, the MMFR test may not eliminate competition effects. Thus, more recovery and more competition would lead to greater PI in the R situation. There are several problems with this interpretation. First, it has often been difficult to demonstrate either spontaneous recovery or specific competition effects in these kinds of verbal situations. Second, the assumption of a direct relationship between recovery and subjective anticipation of a recall test seems to constitute an added complexity in an already overburdened explanatory system. Third, if it is assumed that the MMFR task does not eliminate competition effects, then the concepts of unlearning and competition become, once again, isomorphic and inseparable.

According to the alternative interpretation, PI occurs in the R condition because of an ongoing unlearning process which occurs during the retention interval. Weak items are unlearned in connection with the practice of corresponding other-list strong items. The present data offer some support for this explanation, at least that segment of the explanation which argues that interference effects are related to RIR. First, *S*s in the NR condition were not expecting a recall test and did not display a PI effect. The *S*s in the R situation were expecting a recall test and did demonstrate PI. Second, when *S*s are broken down into high- and low-rehearsal groups, the high-rehearsal group reveals a significant PI effect, while the low-rehearsal *S*s do not. Third, it is interesting to note that the control

R recall is better than the control NR recall, while the experimental R recall is poorer than the experimental NR recall. This suggests that in the control conditions, anticipation of an ensuing recall test helps to maintain or perhaps even strengthens the unopposed associations. On the other hand, the same anticipation of an impending recall test in the experimental or interference conditions leads to a loss of items. The appealing conclusion here is that rehearsal normally occurs in our retention intervals, but that it leads to a relative improvement in control recall where no interfering associations exist and to a decrement in experimental recall through the rehearsal of existing interfering associations.

The general position taken in the present paper, which suggests that PI may be the result of uncontrolled unlearning occurring during the retention interval rather than the result of competition at the time of recall, finds some support in a recent study reported by Ekstrand (1967). Ekstrand found that retention intervals filled with sleep produced spontaneous recovery of first-list associations and a decrement in PI. If PI is determined by competition, one would have expected increased PI in the Ekstrand study. But, quite to the contrary, he found sleep-filled intervals led to a decrement in PI. This result, while inconsistent with the competition interpretation, does make sense if one adopts the notion that PI is determined by RIR and assumes that sleep, as well as the instructions employed in the present study, reduces RIR.

In summary, the data from the present experiment suggest that PI is related to *S*'s anticipation of an impending recall test, which, in turn, is related to amount of RIR. Although the data do not address this question directly, it may be that RIR results in the unlearning of weak second-list associations in the experimental conditions and to the maintenance of the associations in the unopposed control conditions. The exact nature of the unlearning mechanism remains unclear.

REFERENCES

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, 1959, **58**, 97–105.

EKSTRAND, B. R. Effect of sleep on memory. *Journal of Experimental Psychology*, 1967, **75**, 64–72.

HOUSTON, J. P. Proactive inhibition and competition at recall. *Journal of Experimental Psychology*, 1967, **75**, 118–121.

HOUSTON, J. P. Competition in the two-factor theory of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 496–500.

KEPPEL, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior theory and its relation to general S-R theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

KOPPENAAL, R. J. Time changes in the strength of A-B, A-C lists; spontaneous recovery? *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 310–319.

POSTMAN, L., STARK, K., & FRASER, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 672–694.

WIPF, J. L., & WEBB, W. B. Proactive inhibition as a function of the method of reproduction. *Journal of Experimental Psychology*, 1962, **64**, 421.

(Received April 15, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 515–519

SUBJECTIVE RANGE OF THE ODOR OILINESS OF HEPTANOL

KARL E. HENION[1]

University of Texas, Austin

A category, a ratio estimation, and three magnitude scales were constructed from observations of 14 *O*s who judged the oiliness of a geometric dilution series of n-heptanol. Oiliness varied with concentration and had a short subjective range for a fairly typical stimulus concentration range. Ratio scales were stable and resembled power functions with exponents of .184 to .259. Odor oiliness is similar to odor intensity and odor pleasantness in exhibiting a direct relationship between stimulus and subjective ranges. Oiliness does not appear to be prothetic; results supported an earlier suggestion that range size may account for the mixed evidence concerning metatheticity.

The difference in response range of different psychological attributes for different odorants is a relatively unexplored problem of psychophysics (Engen, 1964). Results of direct estimation experiments indicate that the subjective intensity range for a diluted odorant is, typically, relatively small, but that the pleasantness range for a set of undiluted odorants may be quite large (Engen & McBurney, 1964). Much less is known about the range of the oiliness attribute, either for varying concentrations of the same diluted odorant, the subject of this paper, or for different undiluted odorants.

More information about the psychophysics of oiliness is needed, since there is some evidence that smell adaptation may vary with the kind of odorant, possibly its oiliness. Engen (1964), e.g., has indicated that the oiliness of amyl acetate and of heptanal may have contributed to the differential adaptation effect that was found for these odorants in direct scaling experiments of odor quality involving mixtures of the two (Engen, 1964). This indication, he has suggested, may be related to the short falling phase of the electrical potential change in the frog's olfactory organ produced by amyl acetate, contrasted with the longer phase produced by oily odorants, as reported by Ottoson (1956).

The main purpose of the present investigation was to find the range of response to oiliness for various concentrations of heptanol. This odorant is generally recognized to be quite oily and it is the standard for the oiliness attribute used in at least one matching-standards method for characterizing odor qualities (Schutz, 1964). To find a response range for heptanol, two different psychophysical functions were obtained, one by magnitude estimation and one by ratio estimation. Ratio estimation was employed as a control for magnitude estimation and as a way to examine internal consistency. The approach paralleled that of Engen and Lindström (1963), involving the ratio scaling of the odor intensity of amyl acetate. Another question considered was whether oiliness is prothetic or metathetic, and so a category scale was also constructed.

METHOD

Observers.—Fourteen undergraduates, including four women, volunteered to serve without pay as *O*s; none had participated in olfactory experiments before.

Odorants and apparatus.—The *O* sniffed cotton saturated with a given concentration solution of chromatographic-grade heptanol diluted in odorless diethyl phthalate. The cotton used was that wrapped on one end of a Q-tip. The other cotton end was cut off, and this bare end was inserted into the bottom of a small, aluminum-foil-wrapped cork, big enough to be taken up conveniently with the fingers. After each sniff, *O* returned the cork with the attached Q-tip to a 10 × 75 mm. (Pyrex) test tube containing 1.5 cc. of the solution. When the test tube was stopped by the cork, the bottom of the cotton tip was about 1.25 cm. above the surface of the solution.

---

[1] Requests for reprints should be sent to Karl E. Henion, Business Administration - Economics Building 615, University of Texas, Austin, Texas 78712.

There were seven different solutions, consisting of the following concentration series prepared by means of a pipette: 1.5625%, 3.125%, 6.25%, 12.5%, 25%, 50%, and 100% heptanol. The *E* presented the test tubes containing these concentrations singly to *O* on a plastic holder designed for each experimental condition.

Procedure.—The experiments took place in an air-conditioned and ventilated room, where the temperature was maintained at about 25° C. There were five experimental conditions: Three of these required *O*s to make magnitude estimations with different standards, one to make ratio estimations, and one to make category ratings. The *O* was exposed to all conditions, each of which he received in a single 1-hr. session, one condition per day over 5 days, in most cases consecutive days, and in a counterbalanced order.

The *O* was instructed to consider just the oiliness of the stimulus and to ignore all other attributes, like pleasantness or any other qualities, to judge each comparison stimulus independently of another, to sniff consistently, and to place the cotton at the same distance from his nose each time.

The standard for a magnitude estimation condition was 3.125%, 12.5%, or 50% heptanol. The *O* was instructed to let 10 represent the oiliness of the standard, with which he was asked to become familiar. The *E* presented the seven comparison stimuli, including one with the same concentration as the standard, singly and in random sequence with five trials per stimulus. The *O*'s task was to judge the oiliness of each by assigning a number to it proportionate on a ratio scale to the 10 of the standard. He could refer to the standard as often as desired, provided that between sniffs there was at least a pause of 1 min., which was the minimal time interval required between any two sniffs in the experiment. Duplicate sets of comparison stimuli were used to reduce the possibility of recognition of previously presented stimuli from accidental markings on corks or test tubes.

In the ratio estimation condition, *E* presented each of the 21 pairs possible from the seven comparison stimuli once, singly, and in irregular order. Half of the *O*s (every other one) began at one end of the order, and the other half at the other end. The *O*'s task was to sniff the left member of the pair first, then the right one, and to report which was the oilier and what percentage the less oily one was of the more oily one. Each concentration solution appeared equally often on the left as on the right.

In the category estimation condition, there were two standards: *O* was instructed to denote the oiliness of the 1.5625% concentration as 1 and that of the 100% concentration as 5. His task was to assign to a given comparison stimulus an integer from 1 to 5 inclusive which reflected the oiliness of the stimulus in relation to that of the two standards on an equal interval scale. As in each of the magnitude estimation conditions, there were 35 presentations of the seven comparison stimuli, singly presented in random sequence and five trials each. The *O* could refer to the standards as often as desired.

RESULTS

For each magnitude estimation condition, geometric means were computed from the set of 70 magnitude estimates made for every stimulus concentration (14 *O*s, five trials per concentration). The results for the three conditions were combined by adjusting the geometric means for the two conditions with the outer standards to those of the middle standard, i.e., the 12.5% one, and then by computing the geometric means, in turn, from the three proportionalized scales (Engen & Lindström, 1963). This was done in order to get data for a more generalized function (Stevens, 1956), and the results are presented in the first column of Table 1.

In the second column are scale values for the ratio estimation condition. Geometric means were first computed from the set of 14 ratio estimates made for every stimulus pair (14 *O*s, 1 trial/pair) and were arrayed in a ratio matrix. In this form, the data were then analyzed by Ekman's (1958) procedure and the resulting scale values derived. These ratio estimation scale values were also proportionally adjusted to the geo-

TABLE 1

SCALE VALUES OF THE OILINESS OF HEPTANOL BY THREE DIRECT METHODS

Percentage of concentration	Method		
	Magnitude estimation[a]	Ratio estimation[b]	Category rating[c]
1.5625	4.73	.53	1.99
3.125	5.51	.68	2.10
6.25	6.57	.93	2.19
12.5	7.74	1.09	2.87
25	8.97	1.29	3.10
50	9.91	1.37	3.41
100	10.56	1.59	3.89

[a] Adjusted and combined.
[b] Unadjusted.
[c] Arithmetic means.

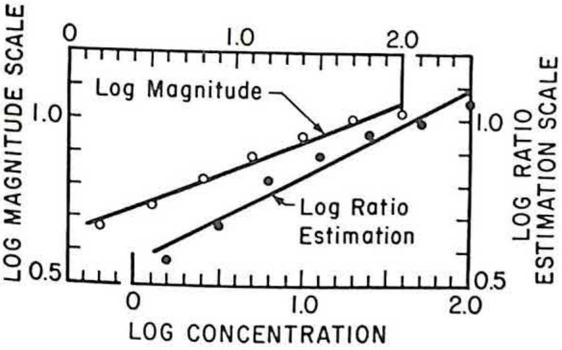


FIG. 1. Log adjusted ratio estimation scale and log combined magnitude estimation scale as functions of log concentrations of heptanol.

metric means of the magnitude estimation condition for the middle standard in order to compare the two kinds of scales.

The comparison is shown in Fig. 1, where the log of the adjusted ratio estimation scale and the log of the combined magnitude estimation scale are plotted as functions of the log concentrations of heptanol. The combined magnitude estimation data fit a straight line well ($r^2 = .98$), and an almost exact fit is prevented by the behavior of the geometric mean at 100% concentration and to a lesser extent by that of the one at 50%. The combined magnitude estimation data conform to a power function with an exponent of .201; the power function property is also evident in the case of the ratio estimation data, which produced an exponent of .259. Again, a linear fit ($r^2 = .96$) would have been almost exact had the possible transducing effect of the 50% and 100% stimulus concentrations been less. From the standpoint of slopes, the two scales compare favorably, although the ratio estimation scale has a somewhat steeper slope than does the combined magnitude scale, a result which was the other way around in the case of subjective intensity for amyl acetate (Engen & Lindström, 1963). Based on the combined magnitude scale, the range of odor oiliness for diluted heptanol is quite short. For the particular largest and smallest stimuli selected in this investigation, the largest sensation was only three times that of the smallest sensation.

Exponents of .210, .207, and .184 were found for the separate magnitude scales whose standards were 3.125%, 12.5%, and 50%, respectively; good linear fits were obtained when the results were plotted for each scale separately (r^2s of .96, .97, and .94, respectively). The exponents are alike to a degree that it is difficult to say much about the influence of the value of the standard on the slope of the function, although the fact that the exponents are in such a narrow range implies that the separate functions are generally stable.

Of the 21 pairs of the ratio estimation condition, there were 6 pairs where the heptanol concentrations of the two members of each pair were in a ratio of 2:1. Similarly, there were 5, 4, 3, 2, and 1 pairs where the stimulus concentration ratios were 4, 8, 16, 32, and 64 to 1, respectively, or altogether a total of six different concentration ratios. Corresponding to these six ratios were the 21 subjective ratios of the previously mentioned matrix. These two sets of ratios were analyzed in a log-log regression plot, shown in Fig. 2. The results indicate that the data are internally consistent generally, in that the least-squares line, to which the data were fitted ($r^2 = .19$), passes near the origin and gives an equation of $R = 1.4S^{.09}$.

Mean category ratings are presented in the last column of Table 1.

DISCUSSION

Like odor intensity, but unlike odor pleasantness, the attribute of odor oiliness appears to have a short subjective range, if the results in

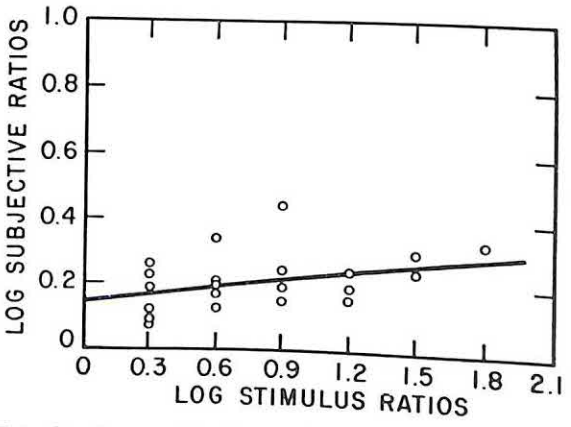


FIG. 2. Log subjective ratios as function of log stimulus ratios of concentrations of heptanol.

the case of diluted heptanol are any indication. In direct methods, it is inappropriate to generalize much about the subjective range of a particular attribute within a modality when the range is expressed as a ratio of the largest to the smallest sensation, since the range depends on the selection of the largest and smallest stimuli. This difficulty is overcome somewhat, e.g., in comparisons of subjective ranges of attributes within the same modality when the stimulus concentration ranges are similar. Thus the oiliness range of heptanol is less than one-half the subjective intensity range that Engen and Lindström (1963) found for amyl acetate over exactly the same stimulus concentration range. But it is about one-fortieth of the hedonic range that Engen and McBurney (1964) found for a representative sample of 18 pleasant and unpleasant undiluted odorants, although comparability involving this attribute is not as appropriate for the reason stated previously. The comparison is mentioned because the range difference is very large. The pleasantness attribute for a diluted odorant varies with its concentration (Moncrieff, 1944); as it now appears, so does the oiliness attribute for diluted heptanol. Although the range of oiliness may differ somewhat for other typically oily odorants, such as hexanol or octanol, one suspects that they may have subjective ranges of the same order of magnitude as that of heptanol over the same stimulus concentration range.

Apparently *O*'s sensitivity to oiliness differences is not very uniform over the continuum, as indicated in Fig. 3, where the category scale is plotted as a function of the magnitude scale.

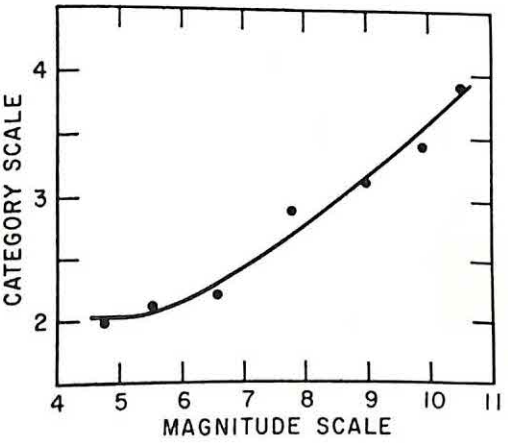


FIG. 3. Category scale as function of combined magnitude estimation scale (fitted by eye).

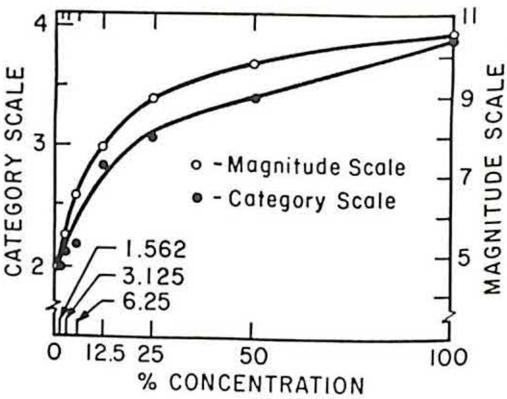


FIG. 4. Category and combined magnitude estimation scales as functions of concentrations of heptanol (fitted by eye).

This result suggests that the attribute may not constitute a metathetic continuum, although linearity is neither a sufficient nor even a necessary condition for establishing the metathetic characteristic. However, the fact that the functional relationship between the category scale and the magnitude scale is not concave downward would tend to rule out oiliness as being prothetic (Stevens & Galanter, 1957); additional evidence in support of this indication is presented in Fig. 4, where both scales are plotted as functions of stimulus concentrations in linear coordinates. Instead of the category scale's being more concave downward than the magnitude scale, which is another acceptable criterion for a prothetic continuum, it is just the other way around.

To pursue the question further, suppose that the physical correlates of the stimuli are not known. Ekman and Künnapas (1963) have argued that category rating methods under this supposition are logarithmic transformations of ratio methods. Such a plot is presented for the data in the present case in Fig. 5, where the category scale is plotted as a function of the log combined magnitude estimation scale, and it is not linear. The curve in Fig. 5 is concave upward, the same shape obtained in a similar plot in the case of pleasantness (Engen & McBurney, 1964) where stimulus correlates were unknown. In both cases, the log of the category scale actually resembles a linear function of the log of the magnitude scale. As already indicated, oiliness has a relatively short subjective range, and it was stimuli having short psychological ranges with which Ekman and Künnapas worked. Engen and McBurney (1964) have suggested that the size of the subjective range may influence the shape of the

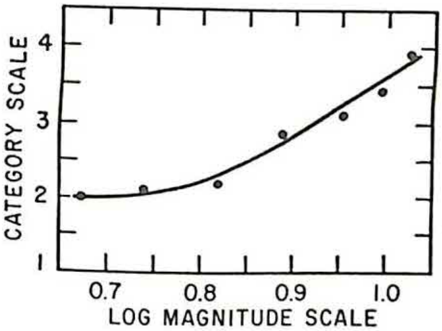


FIG. 5. Category scale as function of log combined magnitude estimation scale (fitted by eye).

function when the two scale types are plotted against each other. If this is so, perhaps there is significance in the two common results obtained for oiliness and pleasantness: (*a*) two different smell attributes having psychological ranges whose sizes for this modality are probably representative of extreme zones, viz., one extremely short and the other extremely long, and (*b*) the two-scale plots for each attribute producing concave upward functions. It would be helpful to know what the two-scale plot looks like in the case of a smell attribute whose subjective range has a more middle-sized value.

The results of the present investigation support the suggestion of Engen and Lindström (1963) concerning the possible uniqueness of olfaction: a direct relationship between effective stimulus range and subjective range and not the usual inverse one as in other modalities. For certain odorants this direct relationship has been exhibited by three odor attributes: intensity (Engen & Lindström, 1963), pleasantness (Engen & McBurney, 1964), and now oiliness.

REFERENCES

EKMAN, G. Two generalized ratio scaling methods. *Journal of Psychology*, 1958, **45**, 287–295.

EKMAN, G., & KÜNNAPAS, T. A further study of direct and indirect scaling methods. *Scandinavian Journal of Psychology*, 1963, 4, 77–80.

ENGEN, T. Psychophysical scaling of odor intensity and quality. *Annals of the New York Academy of Sciences*, 1964, **116**, 504–516.

ENGEN, T., & LINDSTRÖM, C. O. Psychophysical scales of the odor intensity of amyl acetate. *Scandinavian Journal of Psychology*, 1963, **4**, 23–28.

ENGEN, T., & MCBURNEY, D. H. Magnitude and category scales of the pleasantness of odors. *Journal of Experimental Psychology*, 1964, **68**, 435–440.

MONCRIEFF, R. W. *The chemical senses.* New York: Wiley, 1944.

OTTOSON, D. Analysis of the electrical activity of the olfactory epithelium. *Acta Physiologica Scandinavica*, 1965, **35** (Whole No. 122).

SCHUTZ, H. G. A matching standards method for characterizing odor qualities. *Annals of the New York Academy of Sciences*, 1964, **116**, 517–526.

STEVENS, S. S. The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, 1956, **69**, 1–25.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, **54**, 377–411.

(Received April 16, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 520–526

EFFECT OF VARYING CHANNEL CAPACITY ON STIMULUS DETECTION AND DISCRIMINATION[1]

MARILYN C. SMITH[2]

University of Toronto

A decrement in performance is typically found when *S*s must attend to two concurrent stimuli. In order to determine whether this decrement is a function of the available processing capacity, the information associated with one of the tasks (key-press response to a visual stimulus) was manipulated by varying the number of S-R alternatives. Detection or discrimination of concurrently presented auditory tones was then examined, using signal detection procedures, under each of the information conditions. While both detection and discrimination performance were poorer when the subsidiary key-press response was required to the visual stimulus, only the discrimination task was found to be sensitive to the different information conditions. Results were interpreted to support the presence of a central limited-capacity processing mechanism.

It has typically been found that there is a decrement in performance when attention must be paid to more than one stimulus. For example, if *S* is presented with two concurrent stimuli, one auditory and one visual, his absolute judgments of the intensity of the stimuli are far poorer than his judgment about either stimulus presented alone (Lindsay, Cuddy, & Tulving, 1965; Tulving & Lindsay, 1967). As the interval between presentation of the two stimuli is increased, performance tends to improve (Kamlet, 1965). Such experiments are generally interpreted to mean that a person cannot pay attention to two things at the same time and support the notion of a limited-capacity processing mechanism somewhere in the nervous system.

However, such experiments do not make clear the precise location of such a limited-capacity mechanism. For example, the deficit may occur at a rather peripheral sensory stage such that whenever *S* is presented with two signals, he divides his attention equally between them, with the result that neither is perceived as efficiently as when full attention is operative. An example of such a peripheral model is that suggested by Broadbent and Gregory (1963) and by Treisman (1960), which postulates the attenuation of sensory inputs under conditions of multiple stimulation. On the other hand, the limited-capacity channel might occur at a later decision stage where more detailed analysis is made of the information carried by the stimulus. If this were the case, then the greater the information load of a given stimulus, the more channel capacity it would require, causing corresponding decreases in the capacity available for dealing with the second signal.

Previous studies examining the effects of double stimulation on stimulus identification have kept the information content of the two tasks constant, so that it was not possible to separate out these two possibilities. It was the purpose of the present study to examine stimulus identification of one signal under conditions where the information content of a concurrently presented second signal is systematically varied. If performance deficits take place at a peripheral level where a simple division of attention between the two signals occurs, then information content of the stimuli should be unimportant. However, if the limited-capacity mechanism occurs at a later stage, where *S* must make a decision about the stimuli, then increasing the difficulty of one task, thereby increasing the required processing capacity, should result in progressively worsening performance on the second task.

---

[1] Support for this study was provided by National Research Council of Canada Grant APA 191 and Defence Research Board of Canada Grant 9401-38. The author wishes to thank Malcolm Grant for his assistance in running these experiments.

[2] Requests for reprints should be sent to Marilyn C. Smith, Department of Psychology, University of Toronto, Toronto 5, Canada.

In the present experiments, *S* was presented with two concurrent stimuli. In order to avoid possible sensory interference effects, such as masking, stimuli were presented in different modalities, one visual and one auditory. Similarly, to avoid any response interference, a manual key-press response was required to one of the stimuli (visual), while either a detection (Exp. I) or discrimination (Exp. II) judgment was required to the second stimulus (auditory). It was felt that the relatively more simple tasks of detection or discrimination of tones might be more sensitive to changes in attenuation, if this occurred, than a task requiring absolute judgments.

Experiment I examined *S*'s ability to detect a barely audible tone as a function of the difficulty of a manual reaction time (RT) task to a visual stimulus presented simultaneously with the tone. Experiment II examined the influence of the RT task on ability to discriminate the loudness difference of two tones of similar intensity levels. The amount of channel capacity required for the key-press response was systematically varied by manipulating the number of possible S-R alternatives from which a given response was to be selected.

Because difficulty of the visual RT task could influence the criterion level which *S*s adopt in reporting their ability to detect or discriminate between the tones, signal detection procedures were employed. The rating method was used, and the confidence judgments so obtained were used to plot receiver operating characteristic (ROC) curves (Green & Swets, 1966).

EXPERIMENT I

Method

Subjects.—Eight *S*s, all students at the University of Toronto, took part in Exp. I. They were paid $1.50/hr for their services. Six of the *S*s attended three 1-hr. sessions, while the remaining two attended six 1-hr. sessions.

Apparatus.—The apparatus consisted of two consoles, one for *S* and one for *E*, separated by a partition which occluded any visual interaction. The *S*'s console permitted the presentation of both visual and auditory stimuli. The visual stimulus, which consisted of either the number 1 or 2 presented on either a red or green background, was presented via an in-line readout device. The auditory stimulus, a 1,000-cps tone, was presented binaurally through a pair of earphones. White noise could also be presented through the earphones. A small neon bulb, situated 2½ in. above the stimulus window, served as a warning signal. Four push buttons, situated on *S*'s console just below the stimulus window, were used for measuring RT.

The *E*'s console contained a row of buttons for selecting the visual stimulus on each trial and a set of calibrated attenuators for varying the intensity of the tone. An auditory control switch allowed *E* to select either noise alone or noise plus tone on any given trial. RT was measured to the nearest .01 sec. by a Cramer clock situated in *E*'s console. Whenever *S* pressed a given key, it caused a corresponding key to light up on *E*'s console, thereby informing *E* whether a correct response had been made.

The tone was produced by a Hewlett-Packard Model 200 CD wide-range oscillator. The background noise, which had a range of 20 kc., was produced by a General Radio Model 1390-B random noise generator.

Procedure.—The *S*'s ability to detect a barely audible tone was examined as a function of the difficulty of a key-press response required to a visual stimulus presented simultaneously with the tone. The intensity of the tone to be employed was determined for each *S* at the beginning of every session by using the parametric estimation by sequential testing (PEST) method described by Taylor and Creelman (1967). Briefly, this method involved presenting *S* with a series of two consecutive observation periods marked by the onset of a light in the stimulus window. The *S* was presented with white noise continuously through a pair of earphones and a 1,000-cps tone during either the first or second observation period. His task was to decide which of the two observation periods contained the tone and to signal his decision by pressing one of two keys on the panel before him. An arbitrary tone intensity level was modified by successive steps downward when the number of correct responses fell on or above the maximum defined by PEST for a given number of trials and upward when the number correct fell on or below the minimum. The size of each step depended on the size and direction of previous steps and decreased as the test level converged on the target level, a tone intensity level yielding p (correct) $= .75$. The run terminated when a step of .5 db. was called for. The upper and lower limits on correct responses for a given number of trials were determined using a deviation constant of $W = 1.5$ (Taylor & Creelman, 1967).

Each *S*'s ability to detect the tone was tested on 3 consecutive days under each of three conditions: detection alone, detection plus simultaneous two-choice RT task, and detection plus simultaneous four-choice RT task. Order of presentation of conditions was randomized. In all conditions, the intensity of the tone employed was that which gave a 75% correct detection rate as determined at the beginning of each session using the PEST technique.

1. Detection alone. On each of 40 trials, *S* was presented with a 1-sec. warning signal, followed after a 1-sec. foreperiod by a .5-sec. observation period. The duration of the observation period was marked by a light in the stimulus window. On half of the trials, noise plus a 1,000-cps tone was presented during the observation period; white noise alone was presented during the remaining 20 trials. Order of presentation of the tone was randomized. Following each observation period, *S* recorded his decision as to the presence of the tone, together with his degree of certainty, by circling the appropriate response on an accompanying rating scale. A six-point rating scale was employed: (1) yes, very sure; (2) yes; (3) yes, uncertain; (4) no, uncertain; (5) no; (6) no, very sure. After making his decision, *S* pressed a key, indicating he was ready for the next trial.

2. Detection plus two-choice RT task. This condition was similar to the detection-alone condition, except that here the visual stimulus marking the observation period was the stimulus to which *S* was instructed to react as rapidly as possible. This stimulus was either a red or green light, randomly presented. The *S* was instructed to press one key if the light was red and an adjacent one if it was green. Presentation of the tone was randomized and balanced across stimulus conditions. Immediately after reacting to the visual stimulus, *S* recorded his decision regarding the presence or absence of the tone on the six-point scale.

In order to obtain a control measure of the two-choice RT to the visual stimulus presented singly, 24 RT measures were taken to the visual stimulus alone—12 prior to and 12 following the 40 experimental trials.

3. Detection plus four-choice RT task. This condition resembled Cond. 2, except that a four-choice condensation task (Posner, 1964) replaced the two-choice conservation task. A condensation task has been shown to be more difficult than a conservation task (Posner, 1964) and was therefore assumed to require more channel capacity. The visual stimulus marking the observation period was selected from a sample of four, either a 1 or 2, each of which could appear on either a red or green background. Each of the four stimuli (red 1, red 2, green 1, or green 2) was presented 10 times, with a tone accompanying the continuous white noise on 5 of these occasions. The *S* was instructed to press one key if either a red 1 or a green 2 was presented and an adjacent key if either a green 1 or a red 2 occurred. The *S* reacted as quickly as possible to these visual stimuli and then recorded his decision regarding the presence or absence of the tone on the six-point scale.

Once again, 12 control RT measures to the visual stimuli presented alone were taken before and after the double-stimulation trials.

In order to determine whether increased practice would influence the outcome of the experiment, two of the *S*s were run for an additional 3 days, receiving a different random ordering of the three conditions on each day.

Results

The results of the two *S*s who participated in the experiment for an additional 3 days did not differ from the results found for the first 3 days or from those observed for the other *S*s. Consequently, only the data from the first 3 days were included in the analysis.

Detection.—The confidence judgment method was used to obtain ROC curves for the eight *S*s in the manner described by Green and Swets (1966). These curves are shown in Fig. 1. It can be seen that *S*'s ability to detect a tone is superior when he is not required to make a manual response selection to a visual stimulus presented simultaneously with the tone. However, given that there is a decrement in performance as a result of the subsidiary task, the difficulty of this subsidiary task does not appear to influence detectability since detection performance is similar for both RT tasks.

In order to test the significance of the observed differences between conditions, a d' value for each condition, pooled over the 3 days, was obtained for each *S*. This was done by plotting individual ROC curves on normal–normal paper for each condition and taking d' as the point of intersection of the operating characteristic with the negative diagonal (Wickelgren, 1966). In this manner, a d' value was obtained for each condition for each of the eight *S*s, as shown in Table 1. An analysis of variance showed that the difference between conditions was significant at the .01 level, $F\ (2, 14) = 6.59$. To determine which conditions were significantly different from one another, a Duncan new multiple-range test was done. This

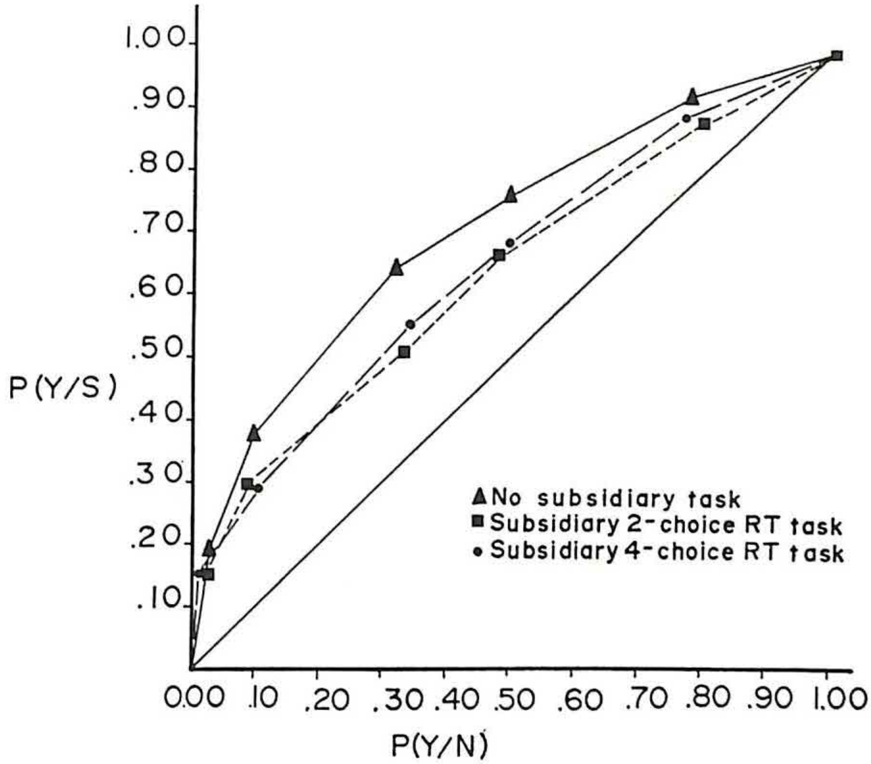


FIG. 1. ROC curves for detection of a 1,000-cps tone presented alone, accompanied by a two-choice conservation task, or accompanied by a four-choice condensation task. (Each curve is the average of eight *S*s.)

test indicated that the detection of the tone presented alone was significantly different ($p < .01$) from detection with either of the subsidiary tasks. However, there was no significant difference in performance as a function of the difficulty of this subsidiary task.

It is also possible to get an approximate idea of the reliability of the difference between two operating characteristics by applying the Kolmogorov-Smirnov two-sample test to the correct detection rates at equal false-alarm rates. Using this test, similar results were found.

Reaction times.—A second indication of the degree of overlap of channel capacity required for the two tasks is provided by an examination of RTs to the visual stimuli when there is no additional task (control measures) and when detection of the tone is required. Mean two-choice RT for the eight *S*s was found to increase from 624 msec. to 926 msec., while mean four-choice RT increased from 967 msec. to 1,169 msec. when the additional task of tone detection was required. A sign test indicated that both these differences were highly significant. This increase, which averaged 250 msec., suggests a decrease in the available processing capacity for the manual key-press response as well as for the detection task.

TABLE 1

d′ VALUES FOR DETECTION OF A 1,000-CPS TONE AS A FUNCTION OF THE DIFFICULTY OF A SUBSIDIARY KEY-PRESS TASK

S	No subsidiary task	Two-choice conservation task	Four-choice condensation task
1	.75	.78	.28
2	.66	−.04	.22
3	1.25	.85	.18
4	.60	.82	.73
5	1.23	.18	.24
6	1.56	.70	1.36
7	1.19	.70	.96
8	.71	−.02	.35
Average	.99	.49	.54

Discussion

The finding that detection is poorer when a second concurrent stimulus must also be attended to provides further generality for the concept of decrement in performance under conditions of double stimulation previously found using more complex measures, such as absolute judgments (Tulving & Lindsay, 1967) or listening tasks (Broadbent & Gregory, 1963). However, it is of interest to note that *S*s' ability to detect the tone was not influenced by the difficulty of the second task. One possible explanation of this result is that the limited-capacity mechanism occurs at a peripheral stage and is determined solely by the requirement of dividing attention between the two signals. Once this division occurs, information content of the signals is unimportant. A second possibility, however, is that a decision about whether or not a tone was presented (i.e., detection) is relatively simple and requires little channel capacity. Hence, the slightly greater reduction in available channel capacity when a four-choice task is presented compared with a two-choice task is not sufficient to cause a significant difference in performance. This explanation would be in agreement with the finding of Kristofferson (1965) that no switch in attention from one modality to another is required for a detection task. However, he has found that such a switch *is* necessary for a discrimination task, suggesting that greater processing capacity is required for discrimination than for detection. Experiment II was therefore done to examine the influence on a discrimination task of changes in available processing capacity.

EXPERIMENT II

Method

Subjects.—Twelve *S*s, chosen from the same pool as in Exp. I, participated in Exp. II.

Apparatus.—The same apparatus was employed as that used in Exp. I.

Procedure.—Each *S* attended two 1-hr. sessions on consecutive days. At the beginning of each session, his 75% auditory difference threshold was determined, once again using the PEST technique, and the difference threshold so determined used for that day. To find this threshold, *S*s were presented with two consecutive pairs of 1,000-cps tones, added to continuous white noise. On each trial, one of the two pairs contained tones of equal intensity, while in the other pair the second tone was slightly louder or softer than the first. The reference tone, presented first in each pair, had an intensity of 73.2 db. The *S* was instructed to press the left key if the tones in the first pair were different and the right key if the tones in the second pair were different. Each tone was presented for .5 sec. The interval between tones was 1 sec., and between pairs there was a 2-sec. pause. Hence, the events of a given trial are as indicated in Fig. 2.

The *S*'s ability to discriminate between the two tones was then examined under each of the three conditions employed in Exp. I: discrimination alone, discrimination plus simultaneous two-choice RT task, and discrimination plus simultaneous four-choice RT task. Each *S* was given a different random ordering of the three conditions on 1 day and this order in reverse on the second day. Control RT measures to the visual stimulus presented alone were taken at the beginning and end of each of the two conditions.

On each experimental trial, the reference tone was presented with the warning signal for .5 sec., followed after a 1-sec. interval by the concurrent presentation of the comparison tone and the visual stimulus for .5 sec. In Cond. 2 and 3, where a key-press response was required to the visual stimulus, *S* reacted as quickly as possible and then indicated on the six-point rating scale whether he thought the two tones were the same or different, together with his degree of confidence. In Cond. 1, no RT response was required. For half of the *S*s the comparison tone was louder than the reference tone on the first

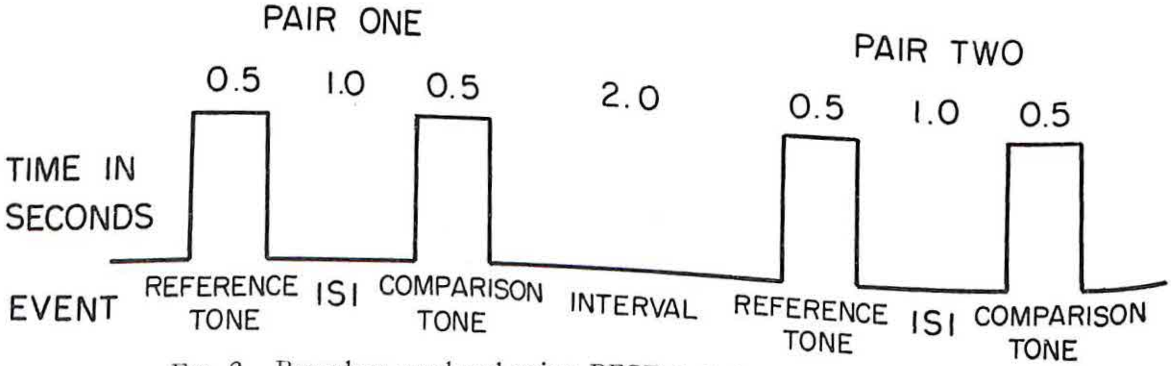


FIG. 2. Procedure employed using PEST technique to determine *S*'s 75% auditory difference threshold.

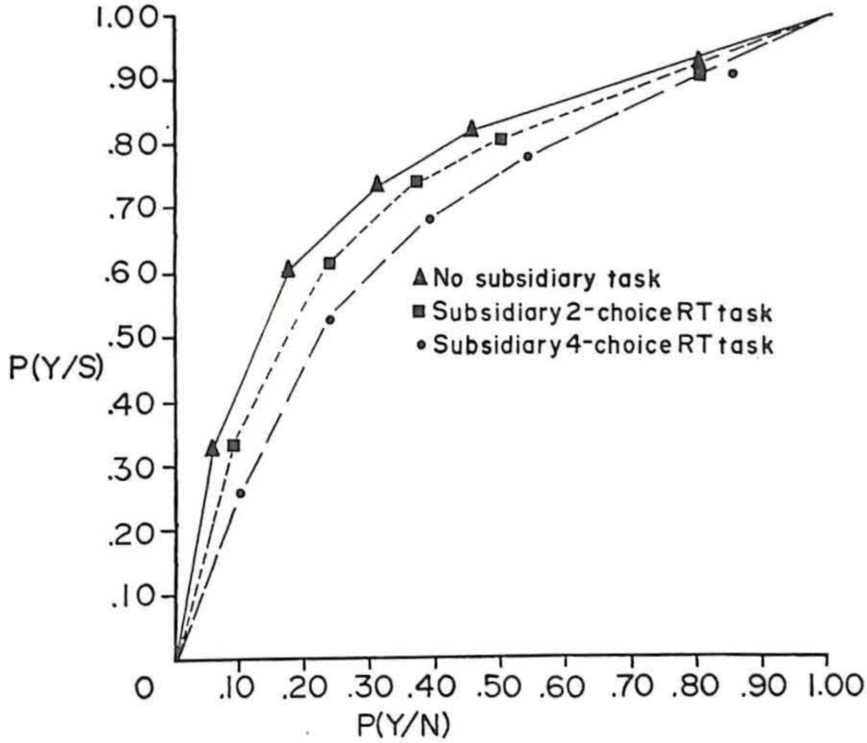


FIG. 3. ROC curves for discrimination between two tones presented alone, accompanied by a two-choice conservation task, or accompanied by a four-choice condensation task. (Each curve is the average of 12 *S*s.)

day and softer than the tone on the second day. For the remaining *S*s this was reversed.

Results

Discrimination.—ROC curves based on the confidence judgment ratings of the 12 *S*s are shown in Fig. 3. It can be seen that *S*s are best able to discriminate between the tones when there is no subsidiary task required. The decrement in performance resulting from the key-press response is greater for the four-choice condensation task than for the two-choice conservation task.

As in Exp. I, to test the significance of these differences, an *F* test was done on the individual d' values for each condition. These values are presented in Table 2. The observed differences between conditions were found to be significant, $F\ (2, 22) = 5.59$, $p < .03$. A Duncan new multiple-range test indicated that all conditions differed significantly from one another, with the exception of the difference between discrimination alone and the two-choice RT condition. Failure to find a significant difference here seems due primarily to *S*4, who discriminated between the tones worst when there was no subsidiary task.

Reaction times.—Two-choice RT was found to increase from 749 msec. to 1,131 msec., while four-choice RT increased from 1,098 msec. to 1,537 msec. when the discrimination task was introduced. This represents

TABLE 2

d' VALUES FOR LOUDNESS DISCRIMINATION OF TWO 1,000-CPS TONES AS A FUNCTION OF THE DIFFICULTY OF A SUBSIDIARY KEY-PRESS TASK

S	No subsidiary task	Two-choice conservation task	Four-choice condensation task
1	.84	.36	.68
2	1.18	1.42	1.05
3	1.92	1.68	.65
4	.60	1.60	.83
5	2.12	1.51	.96
6	.76	.80	.38
7	1.38	.55	.49
8	.42	.05	.66
9	.85	.25	−.13
10	1.43	1.35	1.19
11	2.17	1.77	.78
12	.80	.90	1.10
Average	1.21	1.02	.72

a mean increase of approximately 400 msec. A sign test on the RT data showed that the two differences were highly significant.

Discussion

The results of the present experiment suggest that when *S* must attend to two concurrently presented stimuli, the division of attention is determined by the information content of the signals. Thus as the information load of one of the signals is increased, the capacity available for dealing with the second signal declines, resulting in progressively worsening performance. This finding indicates that the limited-capacity mechanism required to process the two messages is quite centrally located.

Similar results have been reported using a psychological refractory period paradigm, where *S* is required to react to two stimuli in rapid succession (Karlin & Kestenbaum, 1968; Smith, 1969). In these studies it was found that RT to the second signal, whose information was held constant, tended to increase as the information associated with the first signal was increased. Taking the results of the two studies together, it appears that the limited-capacity mechanism probably occurs at a rather central stage, where the information carried by the particular stimuli is analyzed. Further evidence for a central information processor is provided by the work of Posner (1966), who has found increasing decrements in memory as the capacity available for rehearsal is decreased.

One apparently contradictory finding has been reported by Blake and Fox (1969), who recently reported that there was no decrement in visual form recognition threshold following a manual key-press response. However, in their study, a simple, rather than a choice, RT response was required to the first signal (releasing a key to the onset of an auditory tone), and it is possible that this task did not employ a sufficiently large amount of the channel capacity to interfere with the second task. Smith (1967) has found that the psychological refractory period phenomenon does not occur with tasks of high S-R compatibility.

Finally, it might be argued that the worsening performance in the discrimination task is due to the fact that as the RT task becomes more difficult, thereby taking a longer period of time, *S* must wait for increasing periods before giving his decision, thereby allowing increased fading of the memory trace. However, the increase in RT found under the double-stimulation situation suggests that *S* is not merely holding the auditory signal in store, but is actively processing it while making his manual response selection. This is further supported by the relatively greater increase in RT induced by the discrimination task (400 msec.) over that resulting from the detection task (250 msec.).

REFERENCES

Blake, R. R., & Fox, R. Visual form recognition threshold and the psychological refractory period. *Perception and Psychophysics,* 1969, **5**, 46–48.

Broadbent, D. E., & Gregory, M. Division of attention and the decision theory of signal detection. *Proceedings of the Royal Society of London*, Series B, 1963, **158**, 222–231.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Kamlet, A. S. Identification of sequential auditory and visual stimuli. *Psychonomic Science,* 1965, **3**, 419–420.

Karlin, L., & Kestenbaum, R. Effects of number of alternatives on the psychological refractory period. *Quarterly Journal of Experimental Psychology,* 1968, **20**, 167–178.

Kristofferson, A. B. *Attention in time discrimination and reaction time.* (Tech. Rep. CR-194) Washington, D. C.: National Aeronautics and Space Administration, 1965.

Lindsay, P. H., Cuddy, L. L., & Tulving, E. Absolute judgments of simultaneously presented visual and auditory stimuli. *Psychonomic Science,* 1965, **2**, 211–212.

Posner, M. I. Information reduction in the analysis of sequential tasks. *Psychological Review,* 1964, **71**, 491–504.

Posner, M. I. Components of skilled performance. *Science,* 1966, **152**, 1712–1718.

Smith, M. C. S-R compatibility and parallel response selection. *Canadian Journal of Psychology,* 1967, **21**, 496–503.

Smith, M. C. The effect of varying information on the psychological refractory period. *Acta Psychologica,* 1969, **30**, 220–231.

Taylor, M. M., & Creelman, C. D. PEST: Efficient estimates of probability functions. *Journal of the Acoustical Society of America,* 1967, **41**, 782–787.

Treisman, A. M. Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology,* 1960, **12**, 242–248.

Tulving, E., & Lindsay, P. H. Identification of simultaneously presented simple visual and auditory stimuli. *Acta Psychologica,* 1967, **27**, 101–109.

Wickelgren, W. A. Consolidation and retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology,* 1966, **72**, 250–259.

(Received April 16, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 527–531

DIFFERENTIAL CUE HABIT STRENGTH AS A DETERMINANT OF ATTENTION[1]

JOSEPH C. CAMPIONE[2] AND CATHERINE WENTWORTH

University of Washington

An experiment was conducted testing the hypothesis that attention to a dimension in a discrimination learning task is partly determined by the differential habit strengths of the cues along that dimension. The *S*s (64 second-grade children) solved an initial discrimination problem and then were given an extradimensional (ED) shift. On Problem 3, half of the *S*s were given an intradimensional (ID) shift, and remaining *S*s, an ED shift to the dimension relevant on Problem 1. Within each group, for half of the *S*s, the cues employed on Problem 1 were reintroduced, whereas for the remainder, entirely new cues were introduced, yielding a 2 (Shift) × 2 (Cues) factorial design. The differences predicted from the differential cue habit strength assumption were obtained. The results are discussed in terms of a number of attention theories of discriminative learning.

Theorists interested in discrimination learning have come to rely heavily on the concept of selective attention, with various versions of attention theories having been proposed by Atkinson (1958), Sutherland (1959), Kendler and Kendler (1962), Zeaman and House (1963), Lovejoy (1968), and Trabasso and Bower (1968). Basically, it is assumed that the solution of a two-choice discrimination problem requires a chain of two responses: attending to the relevant dimension and making the appropriate choice, or instrumental, response. Concurrent with the rise of attention as a theoretical construct has been an increase in the research devoted to identifying its controlling variables.

The variable which has been most frequently investigated is prior learning, with its effect being illustrated by the greater ease of solving intradimensional (ID) shifts as compared with extradimensional (ED) shifts (Campione, Hyman, & Zeaman, 1965; Dickerson, 1966). In the case of an ID shift, the same dimension is relevant on both the training and transfer problems, whereas for an ED shift, the relevant cues of the shift problem lie along a different dimension from those of the training problem. Other variables which have been identified as determining attention include dimensional preferences (Heal, Bransky, & Mankinen, 1966; Smiley & Weir, 1966), the physical difference between the cues along the dimension (Archer, 1962; Shepp & Zeaman, 1966), and novelty (Trabasso & Bower, 1968).

The inclusion of these latter two variables indicates that *S*s' attention behavior is influenced by the cues presented along the competing dimensions. Apparently, *S*s scan the dimensions and note which cues are available, with the properties of these cues then affecting their attentional response. In addition to physical difference and novelty, one other property of the cues which may be relevant has been suggested by both Lovejoy (1968) in his attention theory and Wyckoff (1952) in his treatment of the role of the (peripheral) observing response in discrimination learning. These authors have assumed that the probability of making an attentional response or observing response is also determined by the differential habit strengths of the cues associated with that response. Thus, e.g., Lovejoy suggested that the probability of attending to a dimension depends in part on the extent to which the cues along that dimension had previously been differentially associated with reward. Here again, *S* is conceived of as looking at the dimensions and noting the cues which are available. For example, the dimensions present might be position (with the associated cues "right" and "left") and color ("red" and "green"). Assume that there is nothing in *S*'s past experience which would dictate a choice of either right or left, but that red has previously been associated with reward, while blue has been associated with nonreward. All other things being equal, the assumption that differential cue habit strength is a determinant of attention would predict that *S* would respond on the basis of color.

---

[1] This study was supported by Grant HD 03917 from the National Institute of Child Health and Human Development. The author would like to thank William R. Shertzer, Assistant Superintendent, Personnel and Research Services, and Jack L. Shore, Principal, and the staff of the Echo Lake School for their cooperation.

[2] Requests for reprints should be sent to Joseph C. Campione, Developmental Psychology Laboratory, University of Washington, Seattle, Washington 98105.

The present experiment, which is summarized in Table 1, was designed to evaluate this assumption. All *S*s originally solved a problem with either color or form relevant (Problem 1) and then, on Problem 2, were given an ED shift to the other visual dimension. The different treatment conditions were introduced on Problem 3, and each *S* was assigned to one of the four major groups. For two of these groups, the stimuli employed in Problem 1 were reintroduced. For one of these groups, the relevant dimension was the dimension which had been irrelevant on Problem 1 and relevant on Problem 2. This group (designated Group ID-old) was thus given an ID shift from Problem 2. For the second group (Group ED-old), the reward contingencies were the same as in Problem 1, and the shift was an ED shift from Problem 2. For the remaining two groups, entirely new stimuli were introduced, and Group ID-new was given an ID shift from Problem 2, with Group ED-new receiving an ED shift from Problem 2. The basic design constituted a 2 (Shift) × 2 (Cues) factorial, and the differential cue habit strength assumption leads to two predictions: (*a*) There should be a Shift × Cues interaction; (*b*) Group ID-new should learn faster than Group ID-old.

While not of primary concern here, the new groups are of interest in themselves, since there have been no experiments with children comparing performance on ID and ED shifts following training on two different dimensions. While the theoretical predictions outlined earlier are clear, extratheoretical considerations suggest the possibility of a different outcome. In the present case, given that *S*s have learned both a color problem and a form problem, we might expect these *S*s to have learned to attend to both dimensions on subsequent problems, thus eliminating the difference between the shift conditions.

METHOD

Subjects.—Sixty-four second-grade children, 32 boys and 32 girls, served as *S*s. They ranged in age from 85 to 107 mo., with a mean of 91.3.

Experimental design.—With sex and dimension being counterbalanced across groups, the overall design was a 2^4 factorial, with four *S*s per cell.

Apparatus.—A modified version of the Wisconsin General Test Apparatus was used. The *S* and *E* sat facing each other with a one-way vision screen interposed. The base of the apparatus contained a revolving circular tray (22-in. diameter) mounted on rollers. Half of the tray contained two wells 5½ in. apart. Before each trial, *E* would place a marble in one of the wells, place the stimuli in position, and revolve the tray, presenting the

TABLE 1

SUMMARY OF THE EXPERIMENTAL DESIGN

Problem	Group			
	ID-old	ED-old	ID-new	ED-new
1		$C_1^+F_1$ C_2F_2 $C_1^+F_2$ C_2F_1		
2		$C_3^+F_3$ C_4F_4 C_3F_4 $C_4^+F_3$		
3	$C_1^+F_1$ C_2F_2 C_1F_2 $C_2^+F_1$	$C_1^+F_1$ C_2F_2 $C_1^+F_2$ C_2F_1	$C_5^+F_5$ C_6F_6 C_5F_6 $C_6^+F_5$	$C_5^+F_5$ C_6F_6 $C_5^+F_6$ C_6F_5

Note.—C and F stand for color and form, with the subscripts denoting specific cues from the dimensions. Not shown are the conditions in which form was relevant on Problem 1 or the other two spatial settings used on each problem.

stimuli to *S*. The entire apparatus, except the one-way screen, was painted gray.

Colored geometric forms printed on 3 × 3 in. white cards served as stimuli. There was a total of 64 color-form patterns resulting from the factorial combination of eight forms (square, circle, cross, triangle, diamond, star, X, and T) and eight colors (red, blue, gray, green, black, brown, orange, and yellow).

Procedure.—Each *S* was brought individually to the experimental room, where he was given instructions indicating that he would be shown two cards and that a marble would be hidden under one of them. Further, *S* was told that if he was careful, he could find the marble every time, and that if he got enough marbles, he could trade them for a toy of his choice at the end of the session.

Training on Problem 1 was begun immediately, a marble serving as the reward for a correct response. Half of the *S*s were given a problem in which color was the relevant dimension, and half were given a problem in which form was relevant. A noncorrection procedure was used throughout the experiment. Position of the positive stimulus was determined by a Gellermann series on all problems, with an independent Gellermann series being used to determine which value of the irrelevant dimension would be paired with the positive stimulus on each trial. The criterion for learning was set at nine correct responses in a block of 10 trials, successive blocks beginning on Trials 1, 6, 11, etc.

To facilitate learning of the first two problems, the special training procedure employed by Dickerson (1966) was used. If *S* made an error on any trial following Trial 15, *E* presented all four stimuli to *S*, the two positive stimuli on one side of the apparatus and the two negative stimuli on the other side. The *E* pointed to the two positive stimuli and told *S* that these two were always correct, then pointed to the negative stimuli and told *S* that they were never correct. No mention was made of either the dimensions or specific cues present. For each *S* who required them, one application of these instructions was sufficient to guarantee learning, and *S*s were then run to the 9/10 criterion. Finally, the assignment of cues to each *S* was dictated by a counterbalancing procedure which guaranteed that within each of the four major groups, each color and each form was positive and negative an equal number of times.

RESULTS

Problems 1 and 2.—As one check on the equality of the four basic groups, the number of *S*s in each group requiring the special instructions on Problem 1, Problem 2, or both was determined. These data were subjected to chi-square analyses, and in no case could the null hypothesis be rejected. In addition, 2^4 factorial analyses of variance were performed on the number of errors to criterion on Problems 1 and 2. On Problem 1, the main effect of shift, $F\ (1, 48) = 4.34$, $p < .05$, and the Shift × Cues interaction, $F\ (1, 48) = 6.92$, $p < .025$, were reliable, both caused by the fact that Group ED-new performed reliably better than each of the other three groups. No reliable effects were obtained on Problem 2.

Problem 3.—The Problem 3 data were such that the assumption of homogeneity of variance was not met. This assumption was met when transformed scores were employed, and the dependent variable in the analysis was $\log_{\text{errors to criterion} + 1}$. A 2^4 factorial analysis of variance was performed on the transformed scores, and the only reliable effect was the Shift × Cues interaction, $F\ (1, 48) = 8.86$, $p < .005$. The geometric mean for Group ID-new was 1.95, for Group ED-new, 7.28; for Group ID-old, 4.63; and for Group ED-old, 3.33. To follow up the interaction, *t* tests (based on the analysis of variance error term) were performed, comparing the ID and ED shifts separately for each cue condition. Where new cues were introduced, the ID shift was learned significantly more quickly than the ED shift, $t(48) = 3.36$, $p < .001$, whereas with the old cues, the difference was nonsignificant ($t < 1$), with the ED shift being learned slightly more quickly.

To test the second prediction derived from the differential cue habit strength assumption, a *t* test was performed comparing Groups ID-new and ID-old. It was predicted that Group ID-new would learn more quickly than Group ID-old, and this difference was obtained, $t(48) = 2.21$, $p < .05$. Lastly, examination of the Problem 3 data failed to reveal any difference in performance between those *S*s who did or did not require help on either or both of the first two problems.

DISCUSSION

These data provide strong support for the assumption that the probability of attending to a dimension is increased if the cues along that dimension have different habit strengths. The first prediction derived from this assumption was that there should be a Shift × Cues interaction. Considering first the new groups, these groups should enter Problem 3 with a high probability of attending to the relevant dimension of Problem 2. Since this dimension is relevant for Group ID-new, but irrelevant for Group ED-new, we should expect Group ID-new to perform better. Considering the old groups, this same argument would apply. However, due to training on Problem 1, differential cue habit strengths should increase the strength of the *relevant* dimension for Group ED-old, while increasing the strength of the *irrelevant* dimension for Group ID-old. Thus, prior learning and differential cue habit strength would be working against each other in this comparison, and we should expect the ID-ED difference obtained with the new cues to be reduced or even reversed with the old cues, generating the Shift × Cues interaction, which was obtained. While the Shift × Cues interaction was also obtained in the analysis of Problem 1, this finding does not affect the interpretation. Even if the groups did differ significantly in their speed of learning, a conclusion which is somewhat questionable given the Problem 2 analysis, this would tend only to make the test of the Problem 3 interaction a conservative one, since the group which performed best on Problem 1 (Group ED-new) was the group which performed most poorly on Problem 3.

Second, considering Groups ID-new and ID-old, both groups should have learned, on Problem 2, to attend to the dimension which was relevant on Problem 3. However, Group ID-old would have differential habit strengths for the cues of the irrelevant dimension, thereby increasing its strength. Since increasing the strength of the irrelevant dimension should retard performance, we would expect Group ID-new to perform better than Group ID-old, which was the obtained result.

Before proceeding, however, it should be pointed out that one alternative explanation of the Shift × Cues interaction can be proposed on the basis of differential instrumental response transfer. Considering Groups ID-old and ED-old, it might be argued that Group ED-old enters the problem with a higher initial probability of making the correct choice response than does Group ID-old. For Group ED-old, the reward contingencies are the same as in Problem 1, so *if* *S*s attend to the relevant dimension, the probability is very high that they will choose the positive stimulus and hence be reinforced, whereas for Group ID-old, the probability of their making the correct choice response given that they attend to the relevant dimension is only .5. Thus, it might be argued that Group ID-old enters Problem 3 with an advantage in the sense of having a higher probability of attending to the relevant dimension and that Group ED-old has an advantage in the sense that it enters Problem 3 with a higher probability of making the correct instrumental response if *S*s attend to the relevant dimension. If these two advantages cancel each other out, the net result would be no difference between the groups, which was the obtained result.

While this source of transfer is clearly present, arguments can be proposed which suggest that it may not be of much import. First, the advantage of Group ED-old could be realized only if *S*s attended to that dimension in the first place. Second, if instrumental learning proceeds quickly, their advantage is not a very great one. Finally, and most importantly, the comparison of Groups ID-old and ID-new in the present experiment is free of this confounding, since both groups enter Problem 3 with the same probability (.5) of making the correct instrumental response. Thus, the finding that Group ID-old learns more slowly than Group ID-new must reflect differences in *S*s' initial probability of attending to the relevant dimension. For Group ID-old, the cues of the irrelevant dimension differ in their habit strengths, thus increasing the probability of *S*s' attending to that dimension and leading to poorer performance.

As mentioned earlier, it seemed possible that the result of training *S*s on two different problems prior to the final problem might eliminate the usual difference obtained in comparisons of ID and ED shifts. The reason for this suggestion was that after having learned a color problem and a form problem, *S*s might, on Problem 3, attend to both dimensions simultaneously. The reliable difference between Groups ID-new and ED-new indicates that this did not happen, although without the appropriate controls it is impossible to determine whether the difference was less than would have been obtained if only one problem had been given prior to test. The suggestion is that the behavior of second-grade children is more compatible with a one-look model than a multiple-look model (cf. Zeaman & House, 1963). Even after having learned both a color problem and a form problem, these *S*s tended to attend to only one dimension on Problem 3.

REFERENCES

Archer, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. *Journal of Experimental Psychology,* 1962, **63,** 616–620.

Atkinson, R. C. A Markov model for discrimination learning. *Psychometrika,* 1958, 23, 309–322.

Campione, J., Hyman, L., & Zeaman, D. Dimensional shifts and reversals in retardate discrimination learning. *Journal of Experimental Child Psychology,* 1965, 2, 255–263.

Dickerson, D. J. Performance of preschool children on three discrimination shifts. *Psychonomic Science,* 1966, **4,** 417-418.

Heal, L. W., Bransky, M. L., & Mankinen, R. L. The role of dimensional preference in reversal and nonreversal shifts of retardates. *Psychonomic Science,* 1966, **6,** 509–510.

Kendler, H. H., & Kendler, T. S. Vertical and horizontal processes in problem solving. *Psychological Review,* 1962, **69,** 1–16.

Lovejoy, E. *Attention in discrimination learning: A point of view and a theory.* San Francisco: Holden-Day, 1968.

Shepp, B. E., & Zeaman, D. Discrimination learning of size and brightness by retardates. *Journal of Comparative and Physiological Psychology,* 1966, **62,** 55–59.

Smiley, S. S., & Weir, M. W. Role of dimensional dominance in reversal and nonreversal shift behavior. *Journal of Experimental Child Psychology,* 1966, **4,** 296–307.

Sutherland, N. S. Stimulus analysing mechanisms. In *Proceedings of a symposium on the mechanization of thought processes.* Vol. 2. London: Her Majesty's Stationery Office, 1959.

Trabasso, T., & Bower, G. H. *Attention in learning: Theory and research.* New York: Wiley, 1968.

Wyckoff, L. B., Jr. The role of observing responses in discrimination learning. Part I. *Psychological Review,* 1952, **59,** 431–442.

Zeaman, D., & House, B. J. An attention theory of retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963.

(Received April 16, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 532–537

WHOLE–PART TRANSFER FROM PAIRED-ASSOCIATE TO FREE RECALL LEARNING[1]

GORDON WOOD[2]

Michigan State University

All *S*s were presented with a paired-associate (PA) list of 18 word pairs and then a free recall (FR) list of 18 words. The control *S*s did not receive any of the FR words in their PA list. Negative transfer occurred when *S*s were given one word from each PA pair for FR. The *S*s who were given either all stimuli or responses from their PA list for FR had less negative transfer than *S*s who were given both stimuli and responses for FR. There was some evidence of positive transfer when *S*s were given half of the PA pairs for FR, especially when the words from each pair were presented consecutively during FR. Although the results were in the predicted direction, the number of PA learning trials (2, 5, 10) did not have a significant influence on transfer.

The results of several studies using a free recall (FR) to FR transfer design (e.g., Tulving, 1966; Tulving & Osler, 1967) and a study using a FR to serial learning transfer design (Wood, 1969b) support the view that higher order memory units are formed during FR learning. That is, *S*s form associations among the words to be recalled during the learning trial, and these associations result in the recall of one word being influenced by the recall or nonrecall of other words. Wood (1969a) used a paired associate (PA) to FR transfer design to investigate further how higher order memory units influence FR learning. All groups received a bidirectional PA list of 18 pairs; the pairs switched direction unsystematically from trial to trial. Each pair of words was considered a higher order memory unit. The number of words in List 2 (0, 6, 12, 18) that belonged to a memory unit formed during List 1 PA learning was manipulated. Only one word from any two-word unit was presented in List 2. If a word from a List 1 pair was presented in List 2, the two-word unit (PA pair) was considered to be inappropriate for List 2 FR learning since recall of the unit would result in one correct and one incorrect response. The amount of negative transfer increased as the number of inappropriate memory units believed to be present increased.

Segal and Mandler (1967) also used a PA to FR transfer design. They had *S*s learn a 16-pair PA list by the anticipation method and then a FR list composed of half of the items from the PA list. Two of the nine groups received a bidirectional PA list in which the pairs switched direction unsystematically from trial to trial. For one group, one word from each PA pair was included in the 16-word FR list. The FR list for the other group consisted of half of the PA pairs presented one word at a time in random order. The *S*s who received the pairs for FR learning were superior to *S*s who received one word from each PA pair, but the difference was not statistically significant. Unfortunately, Segal and Mandler's failure to use a control group (*S*s who did not receive any FR words in their PA list) made it impossible to determine whether negative transfer was obtained for their groups. Wood's (1969a) results suggest that negative transfer is obtained for *S*s who receive a FR list consisting of one word from each PA pair, but it is not clear whether positive or negative transfer is obtained for *S*s who receive a FR list consisting of both words from half of the PA pairs. If higher

---

[1] This investigation was supported by Grant MH 16991 from the National Institute of Mental Health and by a Biomedical Sciences Support Grant from Michigan State University. Thanks are due to Ronald E. Buchter, Lester M. Hyman, Joyce Pennington, and Susan Purcell.

[2] Requests for reprints should be sent to Gordon Wood, Department of Psychology, Olds Hall, Michigan State University, East Lansing, Michigan 48823.

order memory units are formed during FR learning, the prior pairing of the words to be presented for FR should facilitate FR, particularly if the associated words are presented consecutively during the FR learning trials.

The purpose of Exp. I was to investigate transfer from PA to FR learning by manipulating the relationship between the pairs formed during PA learning and the words included in the FR list. The number of trials of PA learning (2, 5, 10) was also manipulated in an effort to determine how degree of PA learning influences subsequent FR learning. The amount of transfer was expected to increase as the degree of List 1 learning increased.

EXPERIMENT I

Method

Design and lists.—A 3 × 4 factorial design was used in which one variable was the number of List 1 trials (2, 5, 10) and the other variable was the List 1 – List 2 relationship. Group 18 received one word from each of the 18 PA List 1 pairs for FR learning. Group 9P-C (paired-constrained) received half of the PA pairs for FR learning; the two words of each PA pair were presented consecutively for the FR learning trials. Group 9P-NC (paired-nonconstrained) also received half of the PA pairs for FR learning, but the two words of each PA pair were not presented consecutively during the FR learning trials. Group 0 (control) did not receive any of the FR words in the PA list.

All 54 words were two-syllable and could be used as nouns, and no word had a frequency of occurrence of less than four per million. There were four PA lists of 18 pairs each and one FR list of 18 words. The words for the FR list were randomly selected from the pool of 54 words. A random presentation order was prepared for the FR list. The successive pairs of two words each were included in the PA list for the 9P-C Group. The FR words were re-paired, and the PA pairs were included in the PA list for Group 9P-NC. Except for the nine pairs not presented for FR, there were no identical pairs for Groups 9P-C and 9P-NC. The PA list for Group 18 was constructed by pairing each word from the FR list with a word not presented for FR. The same 36 words were used to construct the PA lists for Groups 18, 9P-C, and 9P-NC; the groups differed only with respect to the specific pairings. The PA list for Group 0 (control) was prepared by replacing the 18 FR words with 18 words not used for FR.

Procedure.—A total of 120 students from undergraduate courses in psychology at Michigan State University was randomly assigned to the 12 conditions, such that there were 10 *S*s in each condition. Each *S* was instructed to learn both words in each of the 18 List 1 pairs so that they could recall either word when presented with the other. The words were presented for 2, 5, or 10 alternating study and test trials on a Stowe memory drum at a 2-sec. rate. Four different random orders of the pairs were presented for study trials, and each word was presented as a stimulus for half of the test trials. Following List 1 learning, *S*s were given 10 alternating study and test trials on the FR list. The words were presented at a 1-sec. rate, and each *S* received three different presentation orders of the list. After each presentation of the list, the drum was stopped and *S* was given 30 sec. to recall orally the words from the list. Recall was tape recorded.

Results and Discussion

Free recall learning.—The mean number of words correctly recalled for the 10 FR learning trials for the 12 groups of Exp. I is presented in Fig. 1. An analysis of the total number correct for the 10 FR trials revealed that Group 0 was superior to Group 18, F (1, 108) = 5.83, $p < .05$. A separate analysis on the six groups of *S*s who received Cond. 0 and 18 revealed a significant Trials × Conditions (Group 0 vs. Group 18) interaction, F (9, 486) = 10.58, $p < .001$, indicating that the difference between Group 0 and Group 18 increased as the number of FR trials increased. Although Group 9P-C was superior to Group 9P-NC and Group 0, the differences were not significant, Fs (1, 108) = 3.79 and 2.62, $p > .05$, respectively. Thus, there was clear evidence of negative transfer for Group 18, but only weak evidence of positive transfer for Group 9P-C. Furthermore, it was unclear whether the constrained presentation of previously paired words during FR learning resulted in better recall than nonconstrained presentation.

The number of trials of PA learning was expected to have the greatest effect on Groups 18 and 9P-C. Since inappropriate memory units were believed to be established during PA learning for Group 18 and appropriate memory units for Group 9P-C, it was expected that FR performance would decrease for Group 18 and increase for Group 9P-C as degree of List 1 PA learning increased. Although the results were as predicted for *S*s who received 2 and 10 trials of PA learning, the Trials (2 vs. 10) × Conditions (Group 18 vs. Group 9P-C) interaction was not significant, $F\ (1, 108) = 2.02, p > .05$.

The mean number of intrusions over all 10 FR trials was obtained for the 12 groups. Since the number of trials of PA learning did not influence the number of List 2 intrusions, $F\ (2, 108) = 2.60,\ p > .05$, this variable was collapsed. The mean number of intrusions was 12.46, 6.90, 4.13, and 1.20 for Groups 18, 9P-C, 9P-NC and 0, respectively. There were more intrusions for Group 18 than for Group 0, $F\ (1, 108) = 22.75,\ p < .001$, and there were more intrusions for Group 9P-C than for Group 0, $F\ (1, 108) = 5.82,\ p < .05$. Almost all of the List 2 intrusions were words that had been presented in List 1.

Organization.—Bousfield and Bousfield's (1966) measure of intertrial repetitions (ITR) was used to evaluate the extent to which *S*s organized their FR list. The measure is the number of obtained intertrial repetitions or ordered pairs minus the number of pairs expected on the basis of chance. This measure was modified slightly in that an ITR was counted whenever the same two words were recalled consecutively on two successive trials, regardless of the order of recall within the pair. An ITR score was obtained for each of the nine pairs of successive FR trials for *S*s who were given 10 PA trials. The mean number of ITRs for the four conditions is presented in Fig. 2. Group 9P-C was superior to Group 9P-NC, $F\ (1, 36) = 5.49,\ p < .05$, and to Group 0, $F\ (1, 36) = 4.90,\ p < .05$. Group 18 was inferior to Group 0, $F\ (1, 36) = 4.23,\ p < .05$. Thus, the organization data are consistent with the view that either positive or negative transfer can be obtained when using a PA to FR transfer paradigm by manipulating the appropriateness of the associations established during PA learning. Also, the contiguous presentation of previously paired words during FR learning appears to result in higher organization than a noncontiguous presentation order.

The ITR measure may not be the most appropriate measure of the amount of organization if *S*s in the 9P-C and 9P-NC conditions recalled the two words of each pair consecutively, but did not recall the pairs in the same order on successive trials. That is, the ITR measure assesses whether words from the same pair are recalled consecutively and whether words from different pairs are recalled consecutively on successive pairs of trials. A cluster score was counted every time *S*s recalled the words of a PA pair consecutively during FR learning. A maximum cluster score was counted whenever both words of a PA pair were recalled

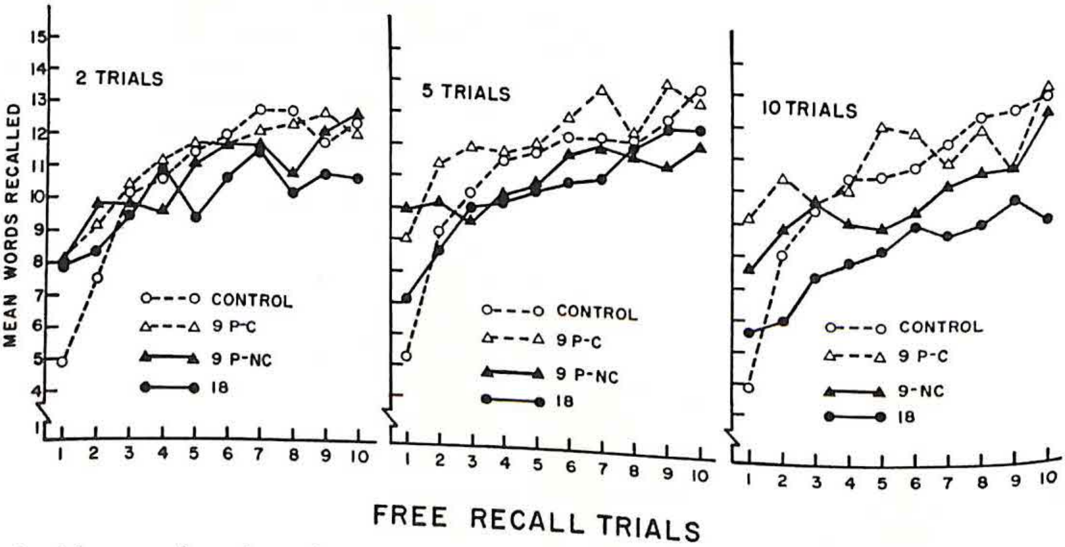


FIG. 1. Mean number of words correctly recalled for the 10 FR trials for the 12 groups of Exp. I.

during FR. When the PA pairs for Group 9P-C were used as the basis for clustering, the maximum number of clusters possible was 573, 345, and 396 for Groups 9P-C, 9P-NC, and 0, respectively. The number of obtained clusters was 95%, 26%, and 32% of the maximum number of possible clusters for Groups 9P-C, 9P-NC, and 0, respectively. When the PA pairs for Group 9P-NC were used as the basis for clustering, Group 9P-NC had a total of 367 clusters out of a maximum possible of 447 (82%). Since all *S*s received the same presentation orders of the FR list, it is clear that the associations established during FR learning were responsible for the large differences in clustering. Although Group 0 and Group 9P-NC did not differ with respect to the ITR measure, the cluster measure revealed that the associations established during PA learning had a strong influence on order of recall for Group 9P-NC. The fact that the cluster measure was higher for Group 9P-C (95%) than for Group 9P-NC (82%) is consistent with the ITR measure in indicating that the presentation order during FR also had an influence on the order of recall.

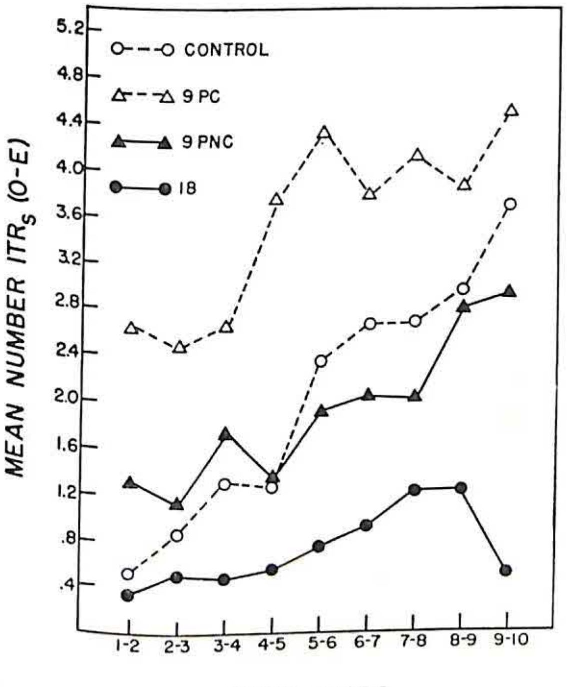


FIG. 2. Mean ITR score (observed minus expected) for the nine pairs of successive FR trials for the four groups in Exp. I who received 10 PA trials.

In summary, the results of Exp. I supported the view that *S*s form higher order memory units during FR learning. There was clear evidence for negative transfer when the memory units established during PA learning were inappropriate for FR learning, and the organization measures indicated that positive transfer is obtained when appropriate memory units are established during PA learning. The positive transfer due to forming appropriate memory units during PA learning appears to be partially dependent upon the consecutive presentation of members of the same memory unit during FR learning. Although the amount of positive and negative transfer seemed to be related to the number of PA learning trials, this variable did not significantly influence transfer.

EXPERIMENT II

Segal and Mandler (1967) found that performance following unidirectional PA learning was higher when the FR lists consisting of one word from each PA pair were constructed by using all the stimuli or responses than when the list contained both stimuli and responses. The group which received half of the PA pairs for FR was about equal to the groups which received either stimuli or responses. Yet, when bidirectional PA learning was used, the subsequent FR learning of the group which received PA pairs was superior to the groups which received one word from each pair. Thus, although Exp. I gave firm support to the view that negative transfer is obtained for *S*s who receive bidirectional PA learning followed by a FR list consisting of one word from each pair, the Segal and Mandler study suggests that positive transfer is obtained when *S*s are given unidirectional PA learning followed by a FR list consisting of either the stimuli or responses from the PA list. The purpose of Exp. II was to assess whether directionality of PA learning influences transfer to FR learning when *S*s are given one word from each PA pair for FR.

Method

Design.—Three groups received a unidirectional PA list of 18 pairs prior to FR learning. Group 18S received the 18 stimuli for FR learning. Group 18R received the 18 responses for FR learning. Group 9S-9R received half of the stimuli and half of the responses for FR learning, such that one word was taken from each PA pair. Group 18 received a bidirectional PA list and one item from each pair for FR learning. The control group (Group 0) did not receive any of the PA words for FR learning. Half of the control *S*s received a bidirectional PA list, and half received a unidirectional PA list.

Lists.—The 18 words for the FR list were randomly selected from the pool of 54 words used in Exp. I. All *S*s received the same FR list. The PA list for Group 18S was prepared by randomly pairing each word from the FR list with one of the 36 remaining words. These same pairs were used for Group 18R and Group 18; however, the stimulus–response order was reversed for Group 18R. The same pairs were also used for Group 9S-9R, but the stimulus–response order was reversed for half of the pairs. The PA list for the control group was constructed by replacing the words presented for FR with the remaining 18 words from the word pool.

Procedure.—A total of 80 students from undergraduate psychology courses at Michigan State University was randomly assigned to the five conditions, such that there were 16 *S*s in each condition. Each *S* in a bidirectional condition was instructed to learn both words in each of the 18 List 1 pairs so that they could recall either word when presented with the other. Each word was presented as a stimulus for half of the test trials. The *S*s in the unidirectional conditions were given standard PA learning trials. The words were presented for 10 alternating study and test trials on a Stowe memory drum at a 2-sec. rate. Each *S* received four different random orders of the PA list. Following List 1 learning, *S*s were given 10 alternating study and test trials on the FR list. The words were presented at a 1-sec. rate, and each *S* received three different presentation orders of the list. After each presentation of the list, the drum was stopped and *S* was given 30 sec. to recall orally the words from the list. Recall was tape recorded.

Results

Free recall learning.—The mean number of words correctly recalled for the 10 FR test trials is presented in Fig. 3. The results were clear in that negative transfer was obtained for all four experimental groups. An orthogonal comparison in which the control group was compared with the average of the four experimental groups revealed a significant difference, F (1, 75) = 38.44, $p < .001$. The orthogonal comparison in which the average of Groups 18 and 9S-9R was compared with the average of Groups 18S and 18R also revealed a significant difference, F (1, 75) = 4.37, $p < .05$. Groups 18S and 18R did not differ significantly, $F < 1$, and Groups 18 and 9S-9R did not differ significantly, $F < 1$.

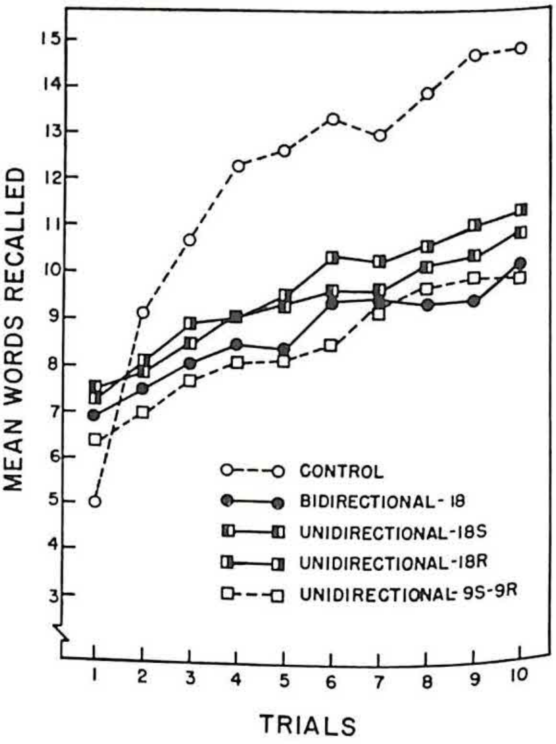


FIG. 3. Mean number of words correctly recalled for the 10 FR trials for the five groups of Exp. II.

The mean number of intrusions during the 10 trials of FR learning was .62, 13.31, 14.44, 17.50, and 18.44 for Groups 0, 18S, 9S-9R, 18R, and 18, respectively. Although the control group had significantly fewer intrusions than the average of the experimental groups, F (1, 75) = 13.27, $p < .001$, there were no significant differences among the experimental groups, $F < 1$. As in Exp. I, almost all of the List 2 intrusions were words that had been presented in List 1.

Organization.—As in Exp. I, an intertrial repetitions (ITR) score was obtained for each of the nine pairs of successive FR trials. The mean number of ITRs for the five conditions is presented in Fig. 4. It is

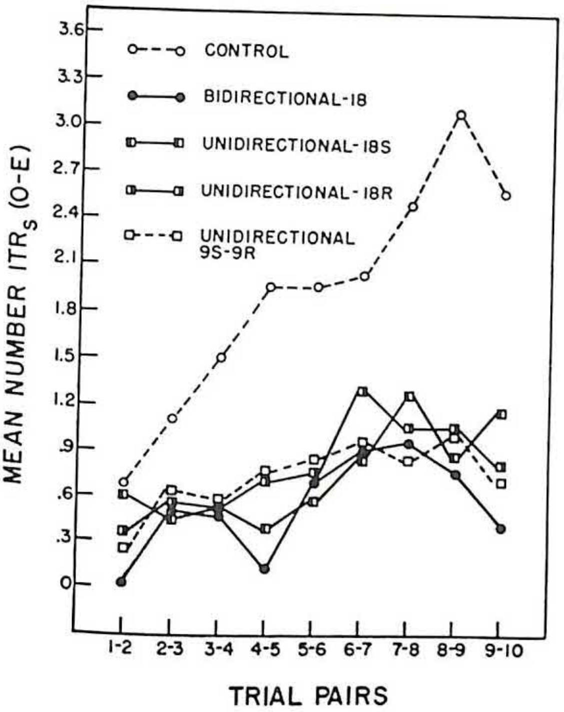


Fig. 4. Mean ITR score (observed minus expected) for the nine pairs of successive FR trials for the five groups of Exp. II.

clear that the control group was superior to the average of the experimental groups, $F\ (1,\ 75) = 14.57,\ p < .001$. Yet, unlike the results obtained for recall, the four experimental groups did not differ significantly, $F < 1$.

Discussion

The results of Exp. II were consistent with the results of Exp. I and with an earlier study by Wood (1969a) in demonstrating that negative transfer is obtained in a PA to FR transfer design when the FR list consists of one word from each PA pair. The directionality of the PA list did not have any discernible effect on FR performance, unless the FR list consisted of all the PA stimuli or responses. The superiority of Groups 18S and 18R to the other two experimental groups supports Segal and Mandler's (1967) contention that *S*s "organize" the stimulus terms and response terms during PA learning. Yet, the fact that marked negative transfer was obtained for Groups 18S and 18R indicates that the positive effect due to the "organization" established among stimuli and responses was considerably less than the negative effect due to the inappropriate associations established between the stimulus and response pairs.

Finally, in a few instances there was a lack of agreement between the organization and recall measures. In Exp. I the organization measures revealed a significant superiority of Group 9P-C over the control group, but the recall measure did not reveal a clear difference between these groups, particularly for the later trials of FR learning. The situation in Exp. II was somewhat different in that the recall measure revealed a significant superiority for the average of Groups 18S and 18R compared to the average for Groups 18 and 9S-9R, but there was no evidence that the experimental groups differed with respect to organization. The failure to find differences in the ITR scores of the experimental groups suggests that Groups 18S and 18R did not form associations among the stimuli or among the response terms. Perhaps, as Segal and Mandler (1967) suggested in an effort to explain similar results, *S*s in Groups 18S and 18R had higher FR performance because their FR words had been coded as "right side" (responses) or "left side" (stimuli) during PA learning. Yet, if the FR list for Groups 18S and 18R had been "organized" during PA learning, these groups should have had fewer intrusions during FR learning. The intrusion data failed to support this prediction.

REFERENCES

Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports,* 1966, **19**, 935–942.

Segal, M. A., & Mandler, G. Directionality and organizational processes in paired-associate learning. *Journal of Experimental Psychology,* 1967, **74**, 305–312.

Tulving, E. Subjective organization and effects of repetition in multitrial free-recall learning. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 193–198.

Tulving, E., & Osler, S. Transfer effects in whole/part free-recall learning. *Canadian Journal of Psychology,* 1967, **21**, 253–262.

Wood, G. Higher order memory units and free-recall learning. *Journal of Experimental Psychology,* 1969, **80**, 286–288. (a)

Wood, G. Whole–part transfer from free recall to serial learning. *Journal of Experimental Psychology,* 1969, **79**, 540–544. (b)

(Received April 23, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 538–544

CONCEPT LEARNING AS A FUNCTION OF THE CONCEPTUAL RULE AND THE AVAILABILITY OF POSITIVE AND NEGATIVE INSTANCES[1]

L. E. BOURNE, JR.,[2] BRUCE R. EKSTRAND, AND BONNIE MONTGOMERY

University of Colorado

One hundred and sixty *S*s, assigned randomly to cells of a 4 (conceptual rule: conjunctive, disjunctive, conditional, and biconditional) × 4 (memory condition: feedback retained for all instances, for positive instances only, for negative instances only, for none of the instances) factorial design, solved an attribute identification problem via the selection paradigm. Rules differed in difficulty and the display of feedback for previously selected instances facilitated performance. The relative effectiveness of retained instances differed, however, with the rule; e.g., positives were more informative in conjunctive and negatives in conditional concepts. There were differences in the stimulus selection patterns of better and poorer *S*s, which also interacted with rule and memory condition.

Positive instances are commonly regarded as a more useful source of information than negative instances in multidimensional concept learning problems (e.g., Hovland, 1952; Hovland & Weiss, 1953). Data to support this contention, however, have been collected in a relatively circumscribed set of problems requiring the identification of a conjunction of relevant attributes. The picture has been clarified and extended by Bourne and Guy (1968), who showed that the relative effectiveness of positive and negative instances depends both on the conceptual rule, i.e., the relationship among relevant attributes of the concept, and on the givens and unknowns of the problem. Specifically, in attribute-identification problems (rule given, relevant attributes unknown), instances of the smaller, more homogeneous type, e.g., negatives of conditional concepts and positives of conjunctive concepts, were used more effectively by *S*s. The outcome was different in rule-learning problems (rule unknown, relevant attributes given), with a mixture of positive and negative instances leading in all cases to more rapid problem solving than either type alone.

The three rules used by Bourne and Guy (1968)—conjunctive, disjunctive, and conditional—plus another, biconditional, which is included in the present experiment, are described and illustrated in Table 1, using the logical bidimensional truth table to show how stimuli are assigned by each of the positive and negative categories. Examples are based on the stimulus population employed in the present study, which consists of 81 patterns generated by four three-valued dimensions. Table 1 also presents the absolute and relative frequencies of positive and negative instances for all rules and the number of stimuli identified with each of the four "truth table" classes.

In the initial comparisons of conceptual rules (e.g., Bourne, 1967; Haygood & Bourne, 1965; Neisser & Weene, 1962), a stable order of rule difficulty emerged. Considering only the bidimensional rules described in Table 1, the order was conjunctive, disjunctive, conditional, and biconditional, from easiest to hardest. This order has been shown to hold both for rule-learning and for attribute-identification problems solved by untrained *S*s (Bourne, 1967) when *S*s are provided with a mixture of positive and negative instances. The data of Bourne and Guy (1968) indicate that rule order

---

[1] This research was undertaken within the Institute for the Study of Intellectual Behavior and is publication No. 3 of the Institute. The work was supported by Research Grant MH 14314 from the National Institute of Mental Health and by a grant-in-aid from the University Council on Research and Creative Work.

[2] Requests for reprints should be sent to Lyle E. Bourne, Jr., Institute for the Study of Intellectual Behavior, University of Colorado, Boulder, Colorado 80302.

TABLE 1

FOUR PRIMARY BIDIMENSIONAL RULES

Relevant attributes		Stimulus classes	Number of stimulus patterns	Truth table description	Rules and category assignments			
Red	Square				Conjunctive[a]	Disjunctive[b]	Conditional[c]	Biconditional[d]
R	S	RS	9	TT	+	+	+	+
R	$\bar{S}$	RTr, RH	18	TF	−	+	−	−
$\bar{R}$	S	YS, BS	18	FT	−	+	+	−
$\bar{R}$	$\bar{S}$	YTr, YH, BTr, BH	36	FF	−	−	+	+

Note.—The following abbreviations are used: T = true (or present); F = false (or absent); R = red; Y = yellow; B = blue; S = square; Tr = triangle; H = hexagon; + = positive instance; − = negative instance; $\bar{x}$ not -x.

[a] Ratio of positive to negative instances = 1:8.

[b] Ratio of positive to negative instances = 5:4.

[c] Ratio of positive to negative instances = 7:2.

[d] Ratio of positive to negative instances = 5:4.

can be changed, and even reversed, when solution-relevant information is provided by one type of instance only. Clearly, the type of instance from which inferences must be drawn is a powerful determiner of conceptual problem solving.

Another generally accepted empirical conclusion is that the availability of previously seen stimulus patterns and their categories facilitates concept learning (Cahill & Hovland, 1960). The opportunity to inspect earlier instances presumably reduces (or eliminates) the memory requirements of the problem, permitting *S* to make inferences directly from physically present information. It is known, moreover, that both stimulus and category must be given, since the availability of the stimulus alone or of a miscategorized stimulus has little or no effect (Pishkin, 1967).

An integration of the data from studies of rule differences, positive and negative instances, and stimulus availability generates a set of empirical questions and certain expectations about their answers. First, the experiments on stimulus availability have been conducted only with conjunctive problems. There is evidence, however, that the more complex the conjunction, i.e., the more relevant attributes conjoined, the more pronounced are the availability effects (Bourne, Goldstein, & Link, 1964). Because conceptual rules are known to differ in difficulty, it might similarly be expected that for a fixed number of relevant attributes, the importance of stimulus availability will increase with rule difficulty.

Second, in the typical experimental procedure, a mixture of positive and negative instances is made available to *S*. It is unlikely, however, that *S*s use the negative stimuli in the typical conjunctive task. Having only negative instances available for inspection might result in no significant improvement in performance, and there might be no difference between the availability of only the positives and of both the positive and the negative instances.

Third, it seems likely that the type of conceptual rule and the type of instance available to *S* will interact. On the basis of results reported by Bourne and Guy (1968), it can be predicted that the availability of positive instances will facilitate solving conjunctive problems, while the availability of negative instances will facilitate solving conditional problems. Furthermore, the availability of both positive and negative instances should, in general, produce better performance than either type alone. Expectations about disjunctive concepts (because the earlier results were indecisive) and biconditional concepts (because they were not included in the earlier study) were uncertain, although both rules partition the stimulus population used in this study into nearly equal halves, which—on the homogeneity of category argument (Bourne & Guy, 1968)—

would suggest a near equivalence of informational content and no reliable difference between positive and negative instances.

The purpose of the present experiment was to investigate these questions and expectations. Independent groups of *S*s were given attribute-identification problems based on four different conceptual rules with all previously seen instances, only positive instances, only negative instances, or no previously seen instances identified and available for inspection. The *S*s selected and tested instances in a self-determined order. In general, stimulus availability was expected to facilitate performance, with the more informative and therefore facilitative instance depending on the particular rule and its characteristic stimulus partition. The proportions of positive and negative instances, fortuitously or purposefully selected by *S*, were also expected to correlate with overall performance.

METHOD

Subjects and design.—The *S*s were 160 student volunteers from introductory psychology classes. They were assigned randomly to 16 primary experimental conditions, 10 *S*s per condition. Experimental conditions were generated by a 4 (rules: conjunctive, disjunctive, conditional, and biconditional) × 4 (availability levels: feedback retained for all instances, for positives only, for negatives only, and no feedback retained) factorial design. Two conceptual problems, differing in their relevant attributes (small-square and one-red), were used, with half of the *S*s within primary conditions assigned to each problem. The *S*s were tested individually in one session lasting 45 min. on the average.

Apparatus.—The stimulus population consisted of the 81 geometrical designs derived from four dimensions (color, form, number, and size), each with three values. Table 1 shows the distribution of these 81 stimuli across truth table classes and, for each rule, across the two response categories. These designs were prepared on 3 × 3 in. cards and were arranged, for simultaneous presentation, on a large upright display panel. Card positions were randomly determined. Response buttons were positioned beneath each card and were used by *S* to make card selections. Also beneath each card were two signal lights, one on either side of the selection button. After each card selection, one light was turned on immediately, the left, green light if the card was positive and the right, red light if it was negative. Depending on the memory condition to which *S* was assigned, the light stayed on either for the duration of *S*'s selection response or for the remainder of the problem. For example, when no instances were to be retained by the apparatus, thus forcing *S* to remember information necessary to solve the problem, signal lights were presented only while the selection button was depressed; when positive instances were to be retained, all green signal lamps once lighted remained lighted, while red lamps were extinguished when *S* released the selection button. The programming of stimulus–signal-light contingencies and the retention of specified signals was accomplished electronically through a network of relays.[3]

TABLE 2

MEAN NUMBER OF CARD SELECTIONS AS A FUNCTION OF THE RULE AND THE AVAILABILITY OF PREVIOUS INSTANCES

Rule	None	Positive only	Negative only	All	Overall mean
Conjunctive	18.5	15.5	15.0	15.0	15.5
Disjunctive	32.0	28.5	26.5	26.5	28.4
Conditional	42.0	33.0	25.5	25.0	31.4
Biconditional	42.0	32.5	45.0	23.0	35.8
Overall mean	33.8	27.4	27.5	22.4	

Task.—In the preliminary instructions, stimulus materials were described and examples of the three values on each of the four stimulus dimensions were shown. The conceptual rule for the problem was explained, using both the truth table (see Table 1) and several illustrative solutions. Finally, *S* was asked to demonstrate his knowledge of the rule by sorting several stimuli into the positive and negative category according to a combination of attributes specified by *E*.

In his experimental problem, *S* selected stimuli at a self-determined rate from the display board by pressing the appropriate buttons. No sample positive instance was shown. After every five selections, *S* was asked to state an hypothesis, i.e., a best guess about the relevant attributes. No feedback as to the correctness of these hypotheses was given; however, the problem was terminated and considered solved when *S* stated the correct hypothesis on two successive occasions.

RESULTS

Mean numbers of card selections before solution for the various primary conditions are presented in Table 2. The analysis of variance in these data revealed significant effects for rules, $F\ (3,\ 128) = 10.09$, $p < .01$, and availability conditions, $F\ (3,\ 128) = 2.89$, $p < .05$. Overall, the order of difficulty was precisely and reliably the same as that reported in earlier work (e.g., Bourne, 1967), $L\ (40) = 1084$, $p < .01$: however, pairwise comparisons on individual degrees of freedom indicate that most of the variance was accounted for by the difference between conjunctive and nonconjunctive concepts, $F\ (1, 128) = 26.62$, $p < .01$. Considering the availability conditions, the most stable difference, statistically, was between some instances and none at all, $F\ (1, 128) = 6.43$, $p < .05$; however, the means show a reasonable progression, with the availability of all instances being numerically superior to either positive or negative instances alone. There are some important internal differences in Table 2, but, in view of the statistically marginal Rule × Availability interaction, $p < .10$, it is difficult to make unambiguous interpretations. In any case, these trends are better exhibited in the following, more detailed analyses.

---

[3] This apparatus was designed and constructed by E. James Archer. The authors are indebted to him for its use in this project.

Selection rates.—At least during the early trials of a problem, chance factors are the main determiners of the information *S* has to work with. Some *S*s selected a large number of positive instances, while others selected predominantly negative instances. The rate of instances of a particular type can be an important consideration for certain rules and availability conditions. To explore this possibility, *S*s within each of the 16 primary conditions (ignoring the problems condition, $F = 1.0$) were divided at the median proportion of positive instances selected into high-rate (HR) and low-rate (LR) groups.

In the analysis of variance, rule and availability effects were again statistically significant, $F\ (3, 128) = 12.37$ and 3.51, respectively, $p < .01$. The overall difference between HR and LR *S*s was unreliable, $F < 1.0$, indicating that the informational content of positive and negative instances cannot be separated from considerations about the rule or about the kind of instances held for *S*'s inspection. The importance of type of instance is revealed by two interactions, Selection Rate × Availability Condition, $F\ (3, 128) = 2.82$, $p < .05$, and Selection Rate × Availability Condition × Rule, $F\ (9, 128) = 2.91$, $p < .01$.

Mean numbers of card selections for HR and LR *S*s within all rule and availability conditions are portrayed in Table 3. Averaging over the four rules, it is of some advantage to select more positive instances when no previous instances are available, but when all instances are available, a higher ratio of negative instances is more useful. These differences were slight, however, and the major interactive effect appears to be attributable to the selection of a relatively high rate of whatever type of instance was retained by the apparatus.

TABLE 3

MEAN NUMBER OF CARD SELECTIONS IN VARIOUS EXPERIMENTAL CONDITIONS FOR *S*s WHOSE SELECTION RATE FOR POSITIVE INSTANCES WAS HIGHER (HR) OR LOWER (LR) THAN THE MEDIAN

Rule	Selection rate	None	Positive only	Negative only	Both
Conjunctive	HR	13	15	11	15
	LR	24	16	15	15
Disjunctive	HR	24	28	26	31
	LR	41	27	27	22
Conditional	HR	55	34	32	26
	LR	29	32	19	24
Biconditional	HR	34	23	67	25
	LR	51	42	23	21
Overall	HR	31.3	25.0	34.0	24.3
	LR	36.3	29.3	21.0	20.5

Interpretation of the two-way interaction must be qualified in terms of the rule defining problem solution. There was a general tendency for *S*s who selected a high proportion of positive instances to perform better on conjunctive problems. The variability among data points was small, however, except when no prior instances of the concept were available; selecting a high proportion of positive instances was clearly facilitative in this case. The *S*s who selected a high rate of positive instances in disjunctive problems were relatively better off when no instances were retained, but worse off when all instances were retained. If only one type of instance was available, no advantage accrued to either type or to either selection rate. The pattern suggests that negative instances are particularly useful when they are retained by the apparatus and can be contrasted with positive instances.

The results for conditional concepts were nearly the mirror image of conjunctions. They showed a clear inverse relation between number of trials to solution and the proportion of negative instances selected. Moreover, the retention of negative instances, alone or in combination with positives, was an important contributing factor to conditional problem solving. Biconditional concepts revealed a pattern similar to that obtained by summing over rules. Performance was better when the type of instance selected with greater frequency was retained. Apparently, *S*s make equally effective use of both positive and negative stimuli in solving biconditional problems, although there might be a slight advantage to the availability of positive instances.

Distribution of stimulus selections.—Which and how many stimulus patterns in the natural population are positive (and negative) instances depend on the solution rule (see Table 1). It is not necessarily the case that all positive (or negative) instances of a particular concept are equally informative. Haygood and Devine (1967), e.g., found that among the positive instances of disjunctive and biconditional concepts, those exhibiting both relevant attributes (TTs) were most useful to *S*. Further clarification of the relation between selection rate and performance, then, might be provided by an analysis of the number of instances from each of the four truth table categories selected by *S* prior to solution. In an effort to determine whether distribution of selections is related to overall success on the task, *S*s within each of the 16 Rule × Availability conditions were divided at the median number of errors to solution into groups of "better" and "poorer" problem solvers.

The effects of rule, $F\ (3, 128) = 25.69$, and availability, $F\ (3, 128) = 7.31$, were statistically significant, $p < .01$, the trends being consistent with those observed in earlier analyses. The better *S*s were, of course, superior to the poorer *S*s, $F\ (1, 128) = 181.54$, $p < .01$, with mean numbers of instances to solution being 16.16 and 38.56, respectively. The overall numbers of selections from the four truth table classes differed reliably, $F\ (3, 384) = 128.54$, $p < .01$, reflecting closely the 1:2:2:4 distribution of the natural stimulus population. Mean selections per class were 3.88, 6.68, 5.89, and 11.33 for the TT, TF, FT, and FF classes, respectively.

TABLE 4

MEAN NUMBER OF CARD SELECTIONS FROM EACH TRUTH TABLE CLASS PRIOR TO PROBLEM SOLUTION

Rule	Truth table class	Group of *S*s	
		Better	Poorer
Conjunctive	TT	2.45	4.20
	TF	2.55	4.75
	FT	2.35	5.00
	FF	2.40	7.30
Disjunctive	TT	1.95	5.05
	TF	4.15	9.90
	FT	4.60	9.05
	FF	7.05	15.00
Conditional	TT	3.05	5.45
	TF	5.60	11.75
	FT	2.40	9.45
	FF	6.45	18.60
Biconditional	TT	3.60	5.30
	TF	3.75	10.95
	FT	4.20	10.05
	FF	8.10	25.70

The more revealing results are contained in the interactions of rule, group, and truth table class. While the overall distribution of selections resembled the natural population, frequencies tended to be highest for the most informative instances of any particular rule, producing a Rule × Truth Table Class interaction, $F\ (9, 384) = 14.14$, $p < .01$. The major examples are TT instances of a conjunction, selected nearly as often as the average of the remaining three categories (which are 2.7 times more prevalent in the population), and TF instances of a conditional concept, selected nearly 1.5 times as often as the equally numerous FTs. As can be seen in Table 4, this general trend is exaggerated in the performance of better *S*s, the Group × Rule × Truth Table Class interaction being highly reliable, $F\ (9, 384) = 5.02$, $p < .01$. Whether by design or by chance, those *S*s who exceeded the median of overall performance encountered a disproportionate number of the more informative instances (or a particular rule) prior to solution. In the sense that better *S*s selected TT instances with relatively high frequency in biconditional problems, the data are consistent with Haygood and Devine (1967); there was no detectable difference in the shape of the distribution of the two groups of *S*s for disjunctive problems, however.

It is clear from the foregoing analyses that the particular stimulus instances, positive or negative, selected by *S* and their distribution over truth table classes are importantly connected with overall performance. Because of the correlational nature of the data, however, it is impossible to decide whether the better *S*s were lucky or, in, fact, intentionally and knowledgeably selected more informative stimuli. One conceivably useful approach to this question is to determine *S*'s pattern of stimulus selections *with respect to his own hypotheses* (rather than with respect to the solution). It seems likely that *S*s with better problem-solving skills will tend to select instances which provide a direct hypothesis test, a TT instance, or which maximize information yield, e.g., a TF instance in a conditional problem. The *S*'s selections on all trials immediately following an hypothesis statement (every fifth trial) were categorized as a TT, TF, FT, or FF instance with respect to the given hypothesis. These frequencies were converted to percentages, *S*-wise, and submitted to analysis of variance.

There were three significant sources of variance in these data. First of all, despite the fact that the percentages of TT, TF, FT, and FF instances available to *S* in the natural population (regardless of his hypothesis) were 11, 22, 22, and 44, respectively, *S* chose 40% TTs, 22% TFs, 21% FTs, and only 17% FFs. There is a sizeable difference among the classes, $F\ (3, 384) = 26.33$, $p < .01$, and a clear preference for the direct test—TT. Choice distributions varied with the rule, however, $F\ (9, 384) = 11.83$, $p < .01$. Instances of the TT class were selected 67% of the time in conjunctive problems, but only 23%, 27%, and 45% of the time in disjunctive, conditional, and biconditional problems. When the concept was conditional, 31% of selections were made from the TF category as compared with 21% FTs (with respect to *S*'s hypothesis).

The Rule × Truth Table Class × Group interaction was marginally significant, $F\ (9, 384) = 1.68$, $p = .06$, indicating some tendency for the better *S*s to choose a higher proportion of informative stimulus instances than the poorer *S*s. There were no other reliable group terms in the analysis, however, and therefore there was no compelling evidence that *S*s with better overall scores are more skillful than poorer *S*s in their problem-solving efforts.

Discussion

These data are generally consistent with and help to extend conclusions drawn in earlier experiments. First of all, they show that a mechanical reduction in the memory requirements of a conceptual problem—by making previous stimuli and their respective categories available for inspection—has a major facilitative effect on performance. The necessity to remember at least some portion of the information provided by preceding events apparently complicates the fundamental inferential requirement of conceptual problems, delays the attainment of a solution, and might induce *S* to adopt a strategy, such as conservative focusing (Bruner, Goodnow, & Austin, 1956), which minimizes the memory requirements of the task. In fact, a focusing analysis, which was not presented, did show a tendency for focusing to increase as the memory requirements increased.

Both the kind of instance selected and the kind retained for *S* impose significant qualifications on the availability effect. In general, the retention of that type of instance selected more often by *S* is the better arrangement. This stands to reason, for it provides the greater amount of logical information, physically present, for *S* to work with. On the other hand, it tends not to be completely consistent with the commonly accepted notion that *S*s make little or no use of negative instances of multidimensional concepts (e.g., Hovland, 1952).

While this overall relationship seems to apply reasonably well to biconditional concepts, it does not hold for the other rule forms. For conjunctives, positive instances are more informative, so that a high selection rate for positive instances is better regardless of other considerations. Just the reverse holds for conditional concepts, in which case it seems to be more important to hold negative instances. These results are consistent with the interpretation of the role of positive and negative instances used by Bourne and Guy (1968), which is based on the relative homogeneity of the two categories. In brief, the argument is that *S*s derive more solution-relevant information in attribute-identification problems from examples of the more limited category, which of course is determined by the rule. In the stimulus population used, positives are fewer in number and more homogeneous for conjunctions, and negatives for conditionals (see Table 1).

Minor complications of this formulation are required for some aspects of performance on disjunctive and biconditional problems. These rules generate nearly equal-sized positive and negative categories of the stimulus population. Overall, the type of instance retained would be expected to make little or no difference in performance. This appears to be the case: *S*s benefit equally from the availability of either instance type. However, for biconditional concepts, retention of the more frequently chosen instance type clearly facilitates performance.

Both the disjunctive and the biconditional rule show a notable contrast between the conditions in which neither and both instance types are retained. When neither is available and *S* must rely exclusively on memory, a high rate of positive instances is better. When both are available, the high negative rate is associated with better performance. This pattern suggests that under ordinary, nonavailability conditions, *S*s adopt a positive focusing strategy, basing hypotheses and inferences primarily on remembered examples of the concept. When instances of both types are retained for *S*, a different and possibly more efficient strategy involving simultaneous contrasts of examples and nonexamples (Wells, 1967) is workable, reducing the relative usefulness of positive instances. There actually appears to be a slight advantage to a higher proportion of negative instances, which is the more homogeneous category under both rules.

As to whether chance or intention determines *S*'s selection of stimulus patterns, the data are suggestive, but not conclusive. The better *S*s do select a more informative distribution of instances. They tend more often than the poorer *S*s to make direct tests of their current hypotheses (by selecting TT or some other critical instance), although the difference between the two groups is only marginally significant. Furthermore, the aforementioned analysis of focusing, which was not presented in the Results section, failed to reveal any impressive differences in focusing between the better and poorer *S*s. On the whole, the distinction between poorer and better *S*s insofar as measures of systematicity in stimulus selections in the present study are concerned is small and unimpressive. There might be more powerful analyses which would give a clearer picture. On the basis of the evidence available here, however, it seems reasonable to conclude that while some *S*s who attained solution with a relatively small number of instances were in fact more skillful problem solvers, chance factors also played a sizeable rôle in the stimulus selections and in the overall quality of performance of all *S*s.

The following statements have generally been accepted as facts about concept learning: (*a*) Positive instances are easier to use or are more informative than negative instances of bidimensional concepts; (*b*) availability of previously seen instance-category contingencies facilitates problem solving; and (*c*) conceptual rules differ in difficulty, in the order conjunctive, disjunctive, conditional, and biconditional. The present experiment imposes certain qualifications on these statements. Plainly, all three variables are systematically interrelated; neither the relative usefulness of positive instances, nor the benefits of stimulus availability, nor the order of rule difficulty is invariant over changes in the other two factors.

REFERENCES

BOURNE, L. E., JR. Learning and utilization of conceptual rules. In B. Kleinmuntz (Ed.), *Memory and the structure of concepts.* New York: Wiley, 1967.

BOURNE, L. E., JR., GOLDSTEIN, S., & LINK, W. E. Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology,* 1964, **67**, 439–448.

BOURNE, L. E., JR., & GUY, D. E. Learning conceptual rules: II. The role of positive and negative instances. *Journal of Experimental Psychology,* 1968, **77**, 488–494.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

CAHILL, H. E., & HOVLAND, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology,* 1960, **59**, 137–144.

HAYGOOD, R. C., & BOURNE, L. E., JR. Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review,* 1965, **72**, 175–195.

HAYGOOD, R. C., & DEVINE, J. V. Effects of composition of the positive category on concept learning. *Journal of Experimental Psychology,* 1967, **74**, 230–240.

HOVLAND, C. I. A "communication analysis" of concept learning. *Psychological Review,* 1952, **59**, 461–472.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology,* 1953, **45**, 165–182.

NEISSER, U., & WEENE, P. Hierarchies in concept attainment. *Journal of Experimental Psychology,* 1962, **64**, 640–645.

PISHKIN, V. Availability of feedback-corrected error instances in concept learning. *Journal of Experimental Psychology,* 1967, **73**, 318–319.

WELLS, H. Facilitation of concept learning by a "simultaneous contrast" procedure. *Psychonomic Science,* 1967, **9**, 609–610.

(Received April 24, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 545–553

ROLE OF SEMANTICS IN REMEMBERING COMPARATIVE SENTENCES[1]

HERBERT H. CLARK[2] AND STUART K. CARD

Carnegie-Mellon University

It is proposed that comparative sentences are decomposed into independent semantic features for remembering. To test this proposal, 64 *S*s were presented with eight types of comparative sentences and were later required to recall them. The sentences were either comparative (*is better than*) or equative (*is as good as*); they were either positive or negative; and they contained either unmarked adjectives (like *good*) or marked ones (like *bad*). In recalling presented sentences, *S*s made systematic errors mostly predictable from semantic feature theory. The *S*s tended to reconstruct unmarked from marked adjectives, affirmations from denials, and sentences meaning "strictly greater than" from sentences meaning "greater than or equal to." These preferences are each accounted for by the loss of one semantic feature in memory, not by the forgetting of the surface structure or syntactic features of a sentence.

Several recent experiments have shown that people retain very abstract semantic information from sentences they hear. Fillenbaum (1966), e.g., found that his *S*s often confused such pairs as *open* and *not closed* in memory. This indicated to Fillenbaum that his *S*s had stored in memory a single semantic "reading" which could be realized as either *open* or *not closed*. Sachs' (1967) *S*s were likewise unable to recognize the particular surface form of a sentence they had recently heard, but were able to recognize its semantic content. Neither Fillenbaum nor Sachs, however, attempted to define what form this abstract representation takes in memory. To this end, Clark and Clark (1968) and Clark and Stafford (1969) proposed that what *S*s retain from sentences are "semantic features" (Bierwisch, 1967; Katz & Fodor, 1963). In many pairs of constructions, one member differs from the other by having one additional semantic feature. The theory predicts that *S*s will often recall the simpler form in place of the more complex one, because the additional feature is easily lost from memory and this causes *S*s to reconstruct the semantically simpler form from the more complex.

One aim of the present study was to extend the findings about semantic features to common English adjectives. The properties of many antonymous adjective pairs, like *good* and *bad*, and *wide* and *narrow*, indicate that such pairs are not symmetrical: that one adjective (*good* and *wide*) is "unmarked," while its opposite (*bad* and *narrow*) is "marked." The evidence for this is twofold: (*a*) Unmarked adjectives can be used with a neutral meaning. For instance, *How good is the news?* usually indicates that the speaker does not expect the news to be either good or bad, whereas *How bad is the news?* implies that he definitely expects it to be bad. And (*b*) unmarked adjectives refer to both a region on a scale and the name of the scale itself. So the name of the *good-bad* scale is *goodness,* not *badness,* and we say *The board is two feet wide,* but not *The board is two feet narrow.*

In terms of the aforementioned theory, an unmarked adjective has one less semantic feature than its marked counterpart; i.e., *bad* consists of all the features of *good* plus the added feature [−*Polar*]. The theory predicts that in comparative sentences the feature [−*Polar*] should be more easily lost in memory than spontaneously added, so that

---

[1] This work was supported in part by United States Public Health Service Research Grant MH-07722 from the National Institute of Mental Health. The authors wish especially to thank Eve V. Clark for her comments and criticisms.

[2] Requests for reprints should be sent to Herbert H. Clark, who is now at Stanford University, Department of Psychology, Stanford, California 94305.

marked adjectives, like *bad,* should be mistakenly recalled as unmarked adjectives, like *good,* more often than the reverse. To test this prediction, the present study required *S*s to recall, after a short period of time, comparative sentences containing marked and unmarked adjectives. (For a more detailed examination and model of marking, see Clark, 1969b; as well as Greenberg, 1966; and Marshall, 1969).

A second aim of the present study was to confirm that *S*s remember comparative sentences generally in terms of semantic components. For this purpose, it is necessary to look closely at comparative (*is better than*) and equative (*is as good as*) constructions. Linguistically, it has been shown (Doherty & Schwartz, 1967; Huddleston, 1967; Lees, 1961; Smith, 1961) that simple comparative and equative sentences are derived from two primitive underlying sentences or base strings. *A is better than B* is derived from *A is good* and *B is good,* two base strings which are conjoined by a series of transformations. *A is better than B* means, literally, that A is good to a greater extent than B is good. It was hypothesized that the component base strings and the relative position of the two terms on a dimension represent semantically distinct components and would therefore be remembered independently of one another; i.e., *S* would remember that A and B are good independently of the fact that A is above B in goodness.

Moreover, the relative position of A and B was itself expected to be remembered as a composite of semantic features. First, both the positive sentence, *A is better than B,* and the negative sentence, *B isn't as good as A,* mean "A is greater than B in goodness." However, both *A is as good as B* and *B isn't better than A* mean "A is greater than *or equal to* B in goodness." (As a shorthand for these two meanings, "greater than" and "greater or equal," respectively, are used even where "less than" and "less than or equal to," respectively, would be appropriate.) Note that the second type of comparison contains all the relational notions of the first type (greater than) plus an additional notion (equal to). It was expected that the additional feature of "equal to" would be easily lost so that "greater or equal" sentences would be reconstructed as their corresponding "greater than" sentences more often than the reverse. Second, it was proposed that positive and negative constructions differ by a single semantic feature. Negative sentences are related to their positive counterparts in that the former deny what the latter affirm. It was hypothesized, therefore, that a denial would be stored in memory as an affirmation plus a feature indicating the sentence was a denial. Since the denial feature should often be lost in memory, negative sentences should be reconstructed as positive more often than the reverse.

METHODS

Materials.—Eight pairs of unmarked-marked antonymous adjectives were selected: *happy-sad, hot-cold, clean-dirty, old-new, pretty-ugly, fast-slow, wide-narrow,* and *good-bad.* Seventy-two triples of nouns, nine triples for each adjective pair, were then composed. Two triples for the adjectives *good* and *bad,* e.g., were *desserts, pie, cake* and *snack, hamburger, hotdog.* The first noun was used as a recall cue and the second two nouns as the two terms in a comparative sentence. For example, some *S*s studied the following noun cue and sentence:

The desserts. The pie isn't as bad as the cake.

Later, presented with only the noun cue, *The desserts,* these *S*s attempted to recall the sentence.

Each triple was used to construct eight comparison sentences. For the adjective pair *good-bad* and the triple (*desserts, pie, cake*), e.g., there were the following eight sentences:

1. The pie is better than the cake.
2. The pie is as good as the cake.
3. The cake isn't better than the pie.
4. The cake isn't as good as the pie.
5. The cake is worse than the pie.
6. The cake is as bad as the pie.
7. The pie isn't worse than the cake.
8. The pie isn't as bad as the cake.

These sentences are distinguished by the following three properties: (*a*) They are either comparative (1, 3, 5, 7) or equative (2, 4, 6, 8); (*b*) they are either positive (1, 2, 5, 6) or negative (3, 4, 7, 8); and (*c*) they contain either unmarked adjectives (1, 2, 3, 4) or marked adjectives (5, 6, 7, 8). The order of *pie* and *cake* was determined at random for the first sentence, but fixed by that decision for the remaining seven sentences. One important property of these eight sentences is that they are all at least partially synonymous. Another is that interchanging the order of the two nouns produces sentences nearly antonymous in meaning to the original eight sentences. There were 72 octuples of sentences altogether.

Procedure.—Every *S* received a study deck and a recall deck. The study deck consisted of the 72 sentences he was to study, 1 from each octuple. Each was printed on a separate IBM card together with its noun cue. The sentences were separated by blank cards into nine blocks of 8 sentences each. The recall deck was precisely identical to the study deck in form and in order except that the sentences to be recalled were omitted.

An *S* studied each sentence in his study deck for 7½ sec., one after another in blocks of 8 sentences. Beginning with the first block, he heard a signal every 7½ sec. to turn from one card to the next. On encountering the blank ninth card, he turned to the recall deck and, working at his own pace, attempted to recall the 8 sentences in order to the appropriate noun cues. Once he had tried 1 sentence and had turned to the next noun cue, he was not allowed to go back. The entire procedure was repeated for the next eight blocks of sentences. The first block was included for practice only and was discarded before analysis, leaving 64 sentences.

The *S*s were 64 introductory psychology students from Carnegie-Mellon University fulfilling a course requirement. They were randomly assigned to eight equal groups, each group receiving one of eight different study decks. The *S*s were run in groups of 8–16 in sessions lasting about 45 min. The *S*s were urged by the instructions to recall the sentences verbatim. When they could not recall a sentence completely, they were to write down what they could remember, even if they could only remember parts of the sentence.

Experimental design.—The study decks were put together according to a carefully counterbalanced design. The goals of the design were to have each *S* encounter one of each adjective pair in each block, each construction in each block, and each adjective pair in each construction. These three specifications were to be uncorrelated with each other as well as with the positions of the sentences within each block. To meet these constraints, eight different study decks were prepared in an 8 × 8 Graeco-Latin square design. First, one order of the 72 noun cues was made for the eight decks; the cues were to be for a different construction type in each deck. The Graeco-Latin square, then, consisted of four orthogonal variables: (*a*) the eight decks, (*b*) the eight adjective pairs, (*c*) the last eight blocks, and (*d*) the eight positions within each block. The practice block in each deck was identical to the very last block, except, of course, that different noun triples had been used. This elaborate counterbalancing was used so that any serial recall strategies or learning-to-learn phenomena would affect all the conditions equally.

RESULTS

In recalling presented sentences, *S*s made many errors. These errors are displayed in Table 1 in a tally of sentence forms *S*s recalled for each presented form; the head of each column lists the form presented, and the head of each row the form recalled. The only responses tallied are those that contained (*a*) either the correct underlying adjective or its opposite, (*b*) a recognizable comparison construction, and (*c*) at least one of the appropriate two nouns. Of the 27% of the responses not tallied, 41% were failures to respond, 25% consisted of nouns with no relational term, and the rest consisted of only an inappropriate relational adjective. In the upper 8 × 8 matrix are tallied the recalled sentences which are partially or wholly synonymous with the presented sentence; in the lower 8 × 8 matrix are those which are partially or wholly antonymous with the presented sentence. In Table 1 and elsewhere, *good* and *bad* represent all unmarked and marked adjectives, respectively.

The errors were statistically analyzed primarily by three-way analyses of variance, with one fixed treatment factor (e.g., the eight construction types) and two random factors (the eight adjective pairs and the 64 *S*s) (Winer, 1962). Adjectives were considered a random factor in order to be able to generalize across adjectives as well as across *S*s (Coleman, 1964). The appropriate error term for testing the fixed factor in such a design is a combination of two-way and three-way interactions, depending on the size of the former. But because calculation of the three-way interactions would have to be carried out on dichotomous values only and because various interaction terms had been systematically confounded, the usual error term could not be used. Instead, the fixed effects were tested against the two random effects separately (adjectives and *S*s), with the reasonable assumption that the Treatment × Adjectives and the Treatment × *S*s interaction variances are zero and their estimates consist solely of error variance.

Parametric F tests were used where possible; nonparametric tests were used for response bias where the scores, though not independent, contained order information appropriate for nonparametric tests.

First, some overall properties of the matrix are discussed, and then two specific properties are dealt with in detail—the effects of the marking of the adjectives and the effects of the type of comparison construction.

Overall results.—The *S*s' preference in recall for certain constructions over others can be seen in the row totals in Table 1. A measure of overall bias towards, say, the form, *X is better than Y,* is obtained by adding the total number of responses listed for Sentence 1 in the top part of the matrix to the total for 1′ in the bottom part. The resulting eight totals differ significantly from each other, χ^2 $(7) = 105$, $p < .001$, Friedman's one-way analysis of variance by *S*s, and χ^2 $(7) = 34.20$, $p < .001$, by adjectives. Among the possible orthogonal comparisons within the eight constructions (an unmarked-marked × positive-negative × comparative-equative design), there were only three significant effects. First, constructions with unmarked adjectives were preferred over those with marked adjectives, 1,635 to 1,347, normal $z = 4.47$, $p < .001$, sign test by *S*s, and $p < .025$, Wilcoxon test by adjectives. Second, positive constructions were preferred over negative, 1,828 to 1,154, normal $z = 6.70$, $p < .001$, sign test by *S*s, and $p < .005$, Wilcoxon test by adjectives. Finally, there was an important interaction between positive-negative and comparative-equative: the positive comparative and negative equative constructions were strongly preferred over the negative comparative and positive equative, 1,726 to 1,256, normal $z = 4.75$, $p <$ .001, sign test by *S*s, and $p < .005$, Wilcoxon test by adjectives. This indicates that the constructions meaning "greater than" (like *is better than* and *isn't as good as*) were preferred over those meaning "greater or equal" (like *is as good as* and *isn't better than*). Notice that this occurs without a significant

TABLE 1

MATRIX OF CONFUSIONS FOR EIGHT LINGUISTIC FORMS

Recalled form	Presented form								Totals
	1	2	3	4	5	6	7	8	
1. A is better than B.	275	46	21	21	4	4	22	34	427
2. A is as good as B.	28	224	5	9	2	4	18	2	292
3. B isn't better than A.	1	4	140	40	7	9	3	1	205
4. B isn't as good as A.	4	3	78	166	21	5	4	5	286
5. B is worse than A.	1	1	13	27	242	45	5	10	344
6. B is as bad as A.	0	0	10	3	13	212	11	6	255
7. A isn't worse than B.	4	3	4	1	2	0	132	36	182
8. A isn't as bad as B.	6	3	0	0	5	4	57	159	234
Subtotals	319	284	271	267	296	283	252	253	2225
1′. B is better than A.	14	7	56	56	15	10	0	1	159
2′. B is as good as A.	4	26	33	32	6	12	2	1	116
3′. A isn't better than B.	17	13	11	1	1	1	14	13	71
4′. A isn't as good as B.	21	17	6	15	0	0	11	9	79
5′. A is worse than B.	6	6	3	4	23	7	51	43	143
6′. A is as bad as B.	1	4	0	0	3	23	24	37	92
7′. B isn't worse than A.	0	1	7	1	10	16	5	3	43
8′. B isn't as bad as A.	2	0	6	7	11	16	5	7	54
Subtotals	65	74	122	116	69	85	112	114	757
Totals	384	358	393	383	365	368	364	367	2982

Note.—*Good* and *bad* are only exemplary adjectives, standing for unmarked and marked adjectives, respectively.

preference for either the comparative or equative syntactic forms themselves.

In parallel with recall bias was *S*s' ability to recall a presented sentence verbatim. Verbatim recall is tallied in the eight main diagonal cells of the upper 8×8 matrix. These eight numbers exhibited the same significant effects that recall bias did. First, constructions with unmarked adjectives were recalled verbatim more often than those with marked adjectives, 805 to 745, $F\ (1, 63) = 4.93$, $p < .05$, by *S*s, and $F\ (1, 7) = 2.00$, $p > .05$, by adjectives. (The latter is not significant presumably because of the insensitivity of the error term with 7 *df*.). Second, positive constructions were recalled more often than negative ones, 953 to 596, $F\ (1, 63) = 91.12$, $p < .001$, by *S*s, and $F\ (1, 7) = 225.6$, $p < .001$, by adjectives. And finally, the constructions meaning "greater than" were better recalled than those meaning "greater or equal," 842 to 708, $F\ (1, 63) = 14.07$, $p < .001$, by *S*s, and $F\ (1, 7) = 23.25$, $p < .001$, by adjectives. There was again no better recall of comparative than of equative constructions.

It might be proposed, however, that the response bias considered previously was an artifact of the better verbatim recall of some constructions over others. This proposition can be tested by considering response bias in Table 1 once verbatim recall, the main diagonal cells of the upper matrix, has been eliminated. Indeed, the "corrected" bias towards the eight constructions shows the same differences as before, since it correlates .85 with verbatim recall. Moreover, the variance of the proportions of the verbatim recall of the eight constructions is only about one-fifth the size of the variance of the proportions of the incorrect recall of those constructions relative to the opportunity for these erorrs. Hence, there is a sizable response bias that is independent of the level of verbatim recall.

Lastly, there appear to be only small differences in *S*s' abilities to recall at least some partly correct sentence for each of the presented sentences. The column totals in Table 1 showed no significant differences by *S*s, although by adjectives there were significant interactions between the markedness of adjectives and positivity, $F\ (1, 7) = 6.63$, $p < .05$, and between markedness and comparative-equative sentence type, $F\ (1, 7) = 6.64$, $p < .05$. It is important that these column totals are approximately equal; otherwise, *S*s might be trying to remember only the easy constructions, and this possibility would considerably complicate the interpretations of response bias and verbatim recall.

TABLE 2

MATRIX OF CONFUSIONS BETWEEN MARKED AND UNMARKED ADJECTIVES

Recalled adjective	Presented adjective		Totals
	Unmarked	Marked	
Unmarked	1394	241	1635
Marked	124	1223	1347
Totals	1518	1464	2982

The lexical marking of adjectives.—A marked adjective like *bad*, we have proposed, should be mistakenly recalled as its unmarked partner *good* more often than the reverse. These two kinds of errors are shown in Table 2, a collapsed version of Table 1 with only two presentation and recall categories remaining. There were 241 errors from *bad* to *good*, but only 124 errors from *good* to *bad*. This difference is consistent across *S*s, normal $z = 4.27$, $p < .001$, sign test, as well as across adjective pairs, $p < .05$, Wilcoxon test. Thus it does appear that adjectives tend to be changed from marked to unmarked over time in memory.

The eight pairs of adjectives, however, were not all alike in their error results. Overall, 66% of the antonym errors were in the predicted direction, but this percentage varied considerably: *pretty-ugly* showed 87%, *happy-sad*, 86%; *old-new*, 80%; *good-bad*, 78%; *fast-slow*, 74%; *hot-cold*, 60%; *wide-narrow*, 54%; and *clean-dirty*, 22%. *Clean-dirty* was the only pair which elicited more errors in the unpredicted than predicted direction.

The adjective pairs also differed considerably in simply how often a presented adjective was mistakenly recalled as its antonym. The overall percentage was 12%, but individual percentages ranged from 19% to 3%: *fast-slow* had 19%, *wide-narrow*, 17%; *clean-dirty*, 15%; *happy-sad*, 13%; *old-new*, 13%; *good-bad*, 13%; *pretty-ugly*, 6%; and *hot-cold*, 3%. An explanation is proposed later to account for the differences in directional antonym errors, as well as for the low percentage of absolute antonym errors for *pretty-ugly* and *hot-cold*.

The four comparative constructions.—There were four types of syntactic constructions in the present study: positive and negative comparative and positive and negative equative. To show the errors among them independent of the adjectives they contained, Table 3 tallies only those responses which preserve the underlying adjective of the presented sentence; it is derived from Table 1 by collapsing across marked and unmarked adjectives. The exemplary adjective throughout this discussion is *good*.

As noted previously, *S*s preferred in recall to give positive over negative constructions and constructions meaning "greater than" over those meaning "greater or equal." These preferences, it is claimed, result from specific errors away from the nonpreferred "greater or equal" and negative constructions.

Consider the positive-negative attribute first. There were 420 negative to positive errors, but only 144 positive to negative ones. That is, 74% of the errors on this dimension were towards the positive.

As for the second kind of errors, there were 424 errors towards "greater than" constructions, but only 242 errors towards "greater or equal" constructions; 65% of the errors were in the predicted direction. These errors, furthermore, were relatively independent of comparative and equative constructions as syntactic forms. When both presented and recalled sentences were positive, 69% of the errors interchanging comparative and equative were towards the comparative, the construction meaning "greater than." But when both were negative, 65% of these errors were towards the *equative*, the construction meaning "greater than." Overall, there were about as many errors towards the comparative as towards the equative, 348 to 308. The *S*s, then, were not making errors based on the syntactic properties of the sentences. Rather, they made errors on a semantic property, recalling "greater than" constructions in place of "greater or equal" ones.

Finally, *S*s preserved the order of the two nouns of the presented sentence in 89% of the recalled sentences. That is, for the presented sentence, *The girl isn't as bad as the boy*, *S*s almost always constructed a sentence with *the girl* and *the boy* placed in the original left-to-right order. Furthermore, they recalled the first noun (the subject of the sentence) better than the second. Of the sentences tallied in Table 1, there were 396 in which either the subject or predi-

TABLE 3

MATRIX OF CONFUSIONS FOR FOUR COMPARISON CONSTRUCTIONS

Recalled form	Presented form				Totals
	I	II	III	IV	
I. A is better than B.	517	91	26	31	665
II. A is as good as B.	41	436	16	15	508
III. B isn't better than A.	3	4	272	76	355
IV. B isn't as good as A.	9	7	135	325	476
Subtotals	570	538	449	447	2004
I′. B is better than A.	37	14	107	99	257
II′. B is as good as A.	7	49	57	69	182
III′. A isn't better than B.	27	29	16	4	76
IV′. A isn't as good as B.	32	33	11	22	98
Subtotals	103	125	191	194	613

cate noun of the presented sentence was missing or wrong. Of these, 63% had the correct subject, and only 37% had correct predicate nouns, a difference significant at $p < .001$, sign test. This percentage was the same whether the noun order had been preserved in recall (63%) or not (62%). That is, it was the subject of the *presented* sentence, not of the *recalled* sentence, that was well remembered.

The recall of noun order appears to be relatively independent of the memory errors towards positive sentences and towards "greater than" sentences. Errors towards the positive occurred 73% of the time when noun order was preserved and 79% of the time when it was not preserved. Similarly, errors towards "greater than" sentences occurred 59% of the time when noun order was preserved and 67% of the time when it was not preserved.

DISCUSSION

It has been argued throughout that remembering sentences is a reconstructive process. A person retains only bits and pieces of what was presented to him and later uses them to reconstruct what he thought was presented. The present proposal is that he retains semantic information. Evidence for this argument shows up on two levels—on the lexical level in the remembering of an unmarked adjective for its opposite and on the syntactic level in the remembering of one comparative construction for another.

Memory for polar adjectives.—In the introduction, comparative sentences like *A is better than B* were shown to consist linguistically of the conjoining, by various transformations, of two base strings, *A is good* and *B is good*. The present *S*s clearly tried to preserve the sense of these primitive constituent sentences, not the meaning of the relation described, for they often recalled *A is better than B* as *A is as good as B*, or as *A isn't as good as B*, or as *B is better than A*. In these instances, *S*s have preserved the underlying base strings while making gross errors in the relation described. At the same time, *S*s almost never recalled *A is better than B* as *B is worse than A* or as *A isn't as bad as B*, two sentences which express the same relation as the presented one, but which contain different base strings. The conclusion is that *S*s stored the base strings—*A is good* and *B is good*—separately from the relation between A and B—that A is higher on the goodness scale than B.

When *S*s did forget the base strings underlying the comparative sentence, they often replaced one adjective by its opposite. These errors were predominantly towards the unmarked adjective, supporting the theory that marked adjectives are stored with one semantic feature in addition to those of its correlative unmarked adjective, and that this feature is sometimes lost in memory. This finding is consonant with several previous studies on free associations (Greenberg, 1966; Marshall, 1968) and on speed of comprehension (Clark, 1969a, 1969b).

The base strings underlying a comparative sentence can be viewed more generally as the *syntactic* presuppositions of that sentence. Underlying *John is sadder than Mary* are the presuppositions that John is sad and Mary is sad. Not all nouns that enter into comparative sentences, however, are neutral vis-à-vis syntactic presuppositions. For example, the sentence. *This genius is stupider than that genius*, sounds highly incongruous, clearly because its presuppositions sound incongruous: *This genius is stupid* and *That genius is stupid*. The word *genius* has its own *lexical* presuppositions—that geniuses are inherently intelligent.

The results show the influence of both syntactic and lexical presuppositions. Syntactic presuppositions, as noted earlier, were well remembered, but lexical presuppositions appeared to have an additional influence on the reconstruction of a sentence at recall. Take, e.g., *clean-dirty*, the one adjective pair which did not show the tendency for errors to go in the direction of the unmarked adjective. Several of the nouns used with *clean* and *dirty* might be more commonly described as dirty than clean: *carpenter, plumber, Ford, Chevy, hand*, or *foot*. For these nouns, an *S* who had forgotten the syntactic presuppositions would reconstruct a sentence having base strings congruent with its lexical presuppositions, i.e., base strings containing *dirty*. This would account for the bias towards *dirty*, the marke form. (Additional evidence for the interaction of lexical and syntactic presupposition appears in the data of Wallis & Audley, 1964 and Audley & Wallis, 1964.) The question remains whether lexical presuppositions migh not account for the more general bias toward the unmarked adjective. The answer appear to be no, for the bias showed up for a sizable cross section of nouns and adjective pairs, and particularly for pairs like *happy-sad, old-new,* and *good-bad* which do not describe lasting or characteristic attributes of objects. The bias towards the unmarked adjective is best accounted for by a simplification of the feature composition of the adjectives.

Actually there is some doubt about whether *hot* of *hot-cold* and *pretty* of *pretty-ugly* are truly unmarked. The dimension for *hot-cold* is really *warmth,* and for *pretty-ugly* it is *beauty;* on the neutrality criterion, too, *hot* and *pretty* appear to be usually nonneutral or marked. These facts are reflected in the number of errors *S*s made on these dimensions. The *S*s interchanged *hot* and *cold* in only 3% of their recalled sentences, and *pretty* and *ugly* in only 6%; these errors were remarkably low compared to the 13–19% range for the other six pairs. Thus in pairs with a clearly unmarked member, antonym errors occurred often, because the two adjectives differ in one feature only and confuse easily. *Hot* and *cold,* and *pretty* and *ugly,* on the other hand, differ in a more complex way, so they did not confuse so easily.

Memory for comparative constructions.—It has been shown that a sentence like *A is worse than B,* when recalled, is reconstructed from two independently retained memories: first, its syntactic presuppositions, that A and B are bad; and second, the comparison of A and B. The comparison relation itself, however, can also be shown to consist of independently remembered features: (*a*) the theme of the sentence, (*b*) whether the sentence was an affirmation or a denial, and (*c*) whether the relation meant "greater than" or "greater or equal."

First, *S*s remembered the surface subject or theme of comparative sentences better than the term in the predicate. Although the theme might be remembered well because it occurs first in these sentences, there is an excellent alternative explanation. Semantically, the theme is the focus of attention in the sentence —what the sentence is about—whereas the other term is not. In remembering sentences, therefore, *S*s could be remembering the term on which the attention of the utterance was focused. There is other evidence which supports this account. For active and passive sentences which differ in thematic emphasis, but not in "cognitive" meaning, the theme is better recalled (Anderson, 1963; Coleman, 1965; Prentice, 1966; Turner & Rommetveit, 1968); theme also has important consequences on the course of problem solving (Duncker, 1945, p. 24), in the generation of sentences (Clark, 1965; Prentice, 1967; Tannenbaum & Williams, 1968), and in the interpretation of sentences (Clark & Begun, 1968; Johnson, 1967; Johnson-Laird, 1968a, 1968b).

Next, *S*s often changed denials into affirmations, and hence remembered affirmations better than denials. This finding supports the hypothesis that the denial feature which marked a sentence as a negative was dropped from memory between study and recall. Such an interpretation is consonant with several previous studies (Gough, 1965, 1966; Wason, 1961; Wason & Jones, 1963).

Finally, *S*s often reconstructed sentences meaning "greater or equal" as ones meaning simply "greater than." This, too, can be viewed as semantic simplification. It occurs because the additional "equal to" feature of "greater or equal" sentences is lost between the presentation of a sentence and its reconstruction.

One last point of interest is a traditional question in memory: How well were the different sentences remembered? The answer, as the present results have shown, depends on the criterion for memory. For the loose criterion of correct recall of one noun and the underlying adjective or its opposite, the eight sentences of the present experiment were about equal. For the stricter criterion of whole or partial synonymy, sentences that were positive or that contained unmarked adjectives were better recalled. For the tightest criterion, that of verbatim recall, sentences that were positive, that contained unmarked adjectives, or that meant "greater than" were better recalled. Each of these levels is accounted for by the simplification of semantic features. Thus, the strength of the present theory is that it accounts for memory at many levels—from remembering the gist of a sentence to recalling the sentence verbatim.

REFERENCES

ANDERSON, B. The short-term retention of active and passive sentences. Unpublished doctoral dissertation, Johns Hopkins University, 1963.

AUDLEY, R. J., & WALLIS, C. P. Response instructions and the speed of relative judgments. I. Some experiments on brightness discrimination. *British Journal of Psychology,* 1964, **55,** 59–73.

BIERWISCH, M. Some semantic universals of German adjectivals. *Foundations of Language,* 1967, 3, 1–36.

CLARK, H. H. Some structural properties of simple active and passive sentences. *Journal of*

Verbal Learning and Verbal Behavior, 1965, **4**, 365–370.

Clark, H. H. Influence of language on solving three-term series problems. *Journal of Experimental Psychology*, 1969, **82**, 205–215. (a)

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76**, 387–404. (b)

Clark, H. H., & Begun, J. S. The use of syntax in understanding sentences. *British Journal of Psychology*, 1968, **59**, 219–229.

Clark, H. H., & Clark, E. V. Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 129–138.

Clark, H. H., & Stafford, R. A. Memory for semantic features in the verb. *Journal of Experimental Psychology*, 1969, **80**, 326–334.

Coleman, E. G. Generalizing to a language population. *Psychological Reports*, 1964, **14**, 219–226.

Coleman, E. B. Learning of prose written in four grammatical transformations. *Journal of Applied Psychology*, 1965, **49**, 332–341.

Doherty, P., & Schwartz, A. The syntax of the compared adjective in English. *Language*, 1967, **43**, 903–936.

Duncker, K. On problem-solving. *Psychological Monographs*, 1945, **58**(5, Whole No. 270).

Fillenbaum, S. Memory for gist: Some relevant variables. *Language and Speech*, 1966, **9**, 217–227.

Gough, P. B. Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 107–111.

Gough, P. B. The verification of sentences: The effects of delay on evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 492–496.

Greenberg, J. H. *Language universals*. The Hague: Mouton, 1966.

Huddleston, R. D. More on the English comparative. *Journal of Linguistics*, 1967, **3**, 91–102.

Johnson, M. G. Syntactic position and rated meaning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 240–246.

Johnson-Laird, P. N. The choice of the passive voice in a communicative task. *British Journal of Psychology*, 1968, **59**, 7–15. (a)

Johnson-Laird, P. N. The interpretation of the passive voice. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 69–73. (b).

Katz, J. J., & Fodor, J. A. The structure of a semantic theory. *Language*, 1963, **39**, 170–210.

Lees, R. B. Grammatical analysis of the English comparative construction. *Word*, 1961, **17**, 171–185.

Marshall, J. Psychological aspects of semantic structure. In R. Meetham (Ed.), *Encyclopedia of linguistics, information and control*. London: Pergamon, 1969.

Prentice, J. L. Response strength of single words as an influence in sentence behavior. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 429–433.

Prentice, J. L. Effects of cuing actor versus cuing object on word order in sentence generation. In, *Studies in language and language behavior*. (Progress Rep. 4) Ann Arbor: University of Michigan Press, 1967.

Sachs, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, **2**, 437–442.

Smith, C. S. A class of complex modifiers. *Language*, 1961, **37**, 342–365.

Tannenbaum, P. H., & Williams, F. Generation of active and passive sentences as a function of subject or object focus. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 246–250.

Turner, E. A., & Rommetveit, R. The effects of focus of attention on storing and retrieving of active and passive voice sentences. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 243–248.

Wallis, C. P., & Audley, R. J. Response instructions and the speed of relative judgments. II. Pitch discrimination. *British Journal of Psychology*, 1964, **55**, 121–132.

Wason, P. C. Response to affirmative and negative binary statements. *British Journal of Psychology*, 1961, **52**, 133–142.

Wason, P. C., & Jones, S. Negatives: Denotation and connotation. *British Journal of Psychology*, 1963, **54**, 299–307.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received April 28, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 554–556

DURATION OF ANTECEDENT DISCRIMINATIVE STIMULI AND WITHIN-SUBJECT REWARD MAGNITUDE DIFFERENCES AS DETERMINERS OF RUNNING SPEED[1]

CARRELL ANN DAMMANN
Emory University

AND

CHARLES C. PERKINS, JR.[2]
Kansas State University

Five groups of eight rats each received a total of 160 runway trials. In each block of 10 trials, 1 trial was to 20 pellets in one color runway (e.g., black) and 9 trials were in the other (white) runway for 5 pellets. Groups differed only in the duration of differential cues before the start-box door was opened. These cues were present for 0, 3, 9, 27, or 81 sec. At asymptote all groups tended to run faster on the more frequent, small-reward trials. This difference decreased as duration of prestart cues increased.

The preparatory response (PR) hypothesis is the hypothesis that classically conditioned responses (CRs) prepare *S* for unconditioned stimulus (UCS) presentation; i.e., following appropriate CRs, the UCS has greater reinforcement value, i.e., is more attractive. This hypothesis, which follows from the assumption that the law of effect applies to classical conditioning as well as to instrumental learning, has been supported by the results of observing response (R_o) experiments in which R_o is acquired in the absence of differential reward (Perkins, 1968).

Recently, Seymann (1969) has applied the PR hypothesis to running speeds in a short runway. His *S*s had nine runs in one color runway (e.g., black) for 5 food pellets (45 mg. each) and one run in the other runway (e.g., white) for a 20-pellet reward. The experimental group had no cues indicating runway color until the start-box door was opened. His advance-cue control *S*s had inserts in the start box which were the same color as the runway on that trial. Both groups rapidly acquired the tendency to run faster on the more frequent small-reward trials than on the less frequent large-reward trials. The difference between large- and small-reward trial speed was greater for the experimental group than for the advance-cue control group. Seymann interpreted his results as indicating that on large-reward trials, *S* had not made appropriate preparatory responses in the start box. Thus *S* ran more slowly to allow more time for preparatory responses to occur while he was in the runway. The control group had 10 sec. of differential cues before the start-box door was opened. During this time, *S* could partially adjust his preparatory responses before the start-box door was opened and thus gained less by an opportunity for further adjustment than did experimental *S*s.

Results for an additional control group indicated that when the same reward magnitude is provided on all trials, but one color runway is used nine times as often as the other, running speeds are the same in the two runways at the end of training. In a second experiment, Seymann (1969) found that with differential reward magnitudes and no differential prestart cues, *S*s ran faster to the more frequent condition and at about the same speed to each when large- and small-reward trials occurred equally often.

The present experiment is a parametric investigation of the effects of duration of differential prestart cues on running speed on small-reward trials minus speed on large-reward trials when only 1 in 10 trials is to a large reward. Such a low proportion of large-reward trials should maximize the tendency of *S*s to slow down on large-reward trials in order to increase PRs. Every attempt was made to keep the procedure comparable to Seymann's (1969).

METHOD

Subjects.—The *S*s were 40 naive, male Sprague-Dawley derived rats, either raised in the Emory

---

[1] The experiment was run as an MA thesis study at Emory University by the first author. The second author was the thesis advisor. The research was supported by National Institutes of Health Grant HD00906 to the second author.

[2] Requests for reprints should be sent to Charles C. Perkins, Jr., Department of Psychology, Kansas State University, Manhattan, Kansas 66502.

University colony or obtained from suppliers. All were 80–108 days old and weighed 224–328 gm. at the beginning of the experiment. Age, weight, and source were counterbalanced for groups.

Apparatus.—The apparatus consisted of a pair of parallel alley runways, one white and the other black. At one end of the runways there was a single gray prestart box, 12 in. long, with two doors opening into the two parallel 18-in. start boxes, which were the same color as their respective runways. At the other end of the 36-in. runway were two 18-in. goal boxes, also corresponding in color to their respective runways. Gray guillotine doors separated the prestart box from the two start boxes, and double guillotine doors were mounted at the entrance to the runways and at the goal boxes. All sections of the apparatus had clear Plexiglas tops. The doors leading from the prestart box were manually operated, but the doors leading from the start box to the runway were operated by release of a catch freeing a counterweight which raised the doors mechanically.

Three Hunter photocells were located at points just beyond the door opening into the runway, the door opening into the goal box, and at the food cup at the end of the goal box. Standard Electric timers were used to measure starting latency (time from opening of the start-box door and triggering the first photocell), running time in the runway (3 ft.), and goal approach time (time between entering the goal box and reaching the food cup 1 ft. away). All measurements were to the nearest 10 msec.

Illumination was provided by a single 75-w. incandescent lamp, 3 ft. to the rear and 3 ft. above the prestart box. The lamp was mounted in a reflector and aimed at the ceiling of the experimental room, so that the bounce light illuminated the entire apparatus fairly evenly. A white noise of 70 db. (SPL) was on whenever *S*s were in the experimental room.

Procedure.—Five groups of eight *S*s each were run in squads containing one *S* in each group. All *S*s were run at 80-85% of their free-feeding body weight and were randomly assigned to one of the five groups. The *S*s were fed once each day at the conclusion of the experimental session, and water was available at all times except when they were in the experimental apparatus. When all *S*s had reached running weight, they were given 1 hr. of apparatus exploration, after which they received 30 free pellets in their home cages.

All *S*s received a total of 160 trials in the apparatus, 5 trials on Days 1 and 2 and 10 trials on each of Days 3–17 inclusive. The only difference between groups was the length of delay time in the start box, i.e., duration of exposure to discriminative cues.

On each trial, *S*s were placed in the gray prestart box for 10 sec. At the end of this period, one of the two doors opened, exposing *S* to either a black or a white start box. The door to the start box was closed behind *S* upon his entry. The *S*s in Groups 0, 3, 9, 27, and 81 were delayed in the start box for 0, 3, 9, 27, or 81 sec., respectively. For *S*s in Group 0, the door leading into the runway was open at all times. For other groups, the door into the runway was opened at the end of the delay period. The goal box contained either 5 or 20 45-mg. Noyes pellets. On 9 of the 10 trials in a block (e.g., on 1 day), the start box, alley, and goal box were one color (e.g., black) and the reward magnitude was 5 pellets. On 1 of the trials (the test trial), the start box, alley, and goal box were the opposite color (e.g., white) and the reward magnitude was 20 pellets.

The 20-pellet trials came in each ordinal position from 3 to 10 on 2 of the 16 10-trial blocks. The ordinal position of the test trial was unpredictable and was never the same for two *S*s in a group on the same day.

Color of the 20-pellet runway was black for three squads and white for three. In the remaining two squads, three *S*s were run on one color-reward combination and two *S*s on the other to check for possible effects of greater use within a squad of one alley than another. Each of the five *S*s was run on 1 trial before the first *S* received his next trial. After the first 20 trials, the intertrial interval was fairly stable at about 8 min.

Measures of runway time and goal-box time were recorded for each *S* on each trial. Measures of starting latency were available for all *S*s except those in Group 0.

RESULTS AND DISCUSSION

As a check on possible effects of predominant use of one alley within a squad, two of the eight squads were run with the color of the more frequently used (five-pellet) runway alternated for successive *S*s in the squad. Since the results for these squads did not differ systematically from those of others, predominant use of one alley cannot explain the results reported subsequently.

To provide measures similar to Seymann's (1969), all measures were converted to running speeds by taking reciprocals of time scores for each measure on each trial. Measures on the first two trials of each day have been omitted from all data analyses to eliminate variability resulting from warm-up effects and to keep measures comparable to Seymann's. Such warm-up effects were apparent in the present experiment as in Seymann's.

Figure 1 shows mean speed in ft/sec (4/running time) for runway and goal box combined on large- and small-reward trials separately for the last 3 days. It will be noted that there is a systematic tendency for

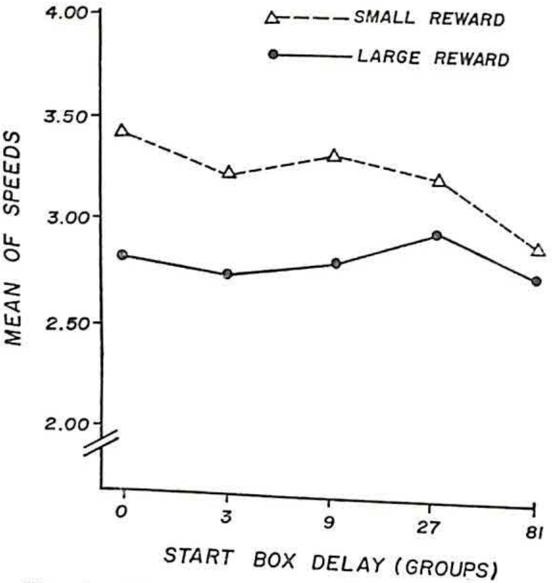


FIG. 1. Mean speed for runway and goal box combined for each group on the last 3 days of training. (See text for details.)

the differences between large- and small-reward trials to decrease as start-box delay time increases. The differences would be ordered in the predicted way due to chance alone only 1 time in 120. A simple analysis of variance was also significant, $F\ (4, 35) = 4.02$, $p < .01$. Thus the predicted relationship was obtained.

Trends were similar but less reliable for runway and goal-box measures separately. Analysis of variance of runway speed differences was significant, $F\ (4, 35) = 3.60$, $p < .05$, but goal-box speeds were not, $F\ (4, 35) = 1.51$, $p > .10$.

Speed differences have been emphasized, rather than separate speed measures on large- and small-reward trials, because difference measures are less affected by individual differences in running speed (Perkins, 1970). Thus they are a great deal more sensitive. Nevertheless, the theoretical interpretation of the role of PRs is that they decrease running speed to the large reward rather than increase speed to small reward. Examination of Fig. 1 indicates more of a tendency for small-reward speeds to be a negative function of cue duration than for large-reward speeds to be a positive function. However, a simple analysis of variance of mean speeds for large-reward trials on the last 3 days combined indicates that the differences are not significant, $F\ (4, 35) = 1.59$, $p > .10$. Only the difference measures were sensitive enough to pick up a significant effect.

The failure of *S*s in even Group 81 to run faster to large reward than to small reward suggests that 81 sec. may not be long enough for a change in the PRs to be complete. Valle (1968) has reported significantly more eating with 60- than with 15-min. exposure to feeding-related cues, a finding which suggests that some component PRs may take longer than 15 min.

Further work is needed to determine if start-box delay affects running speed for *S*s receiving only one reward magnitude. Also, the effect of different proportions of large- and small-reward trials when there are differential prestart cues demands further study.

The results make it quite clear that an important variable in within-*S* reward magnitude experiments is prestart differential cues. It is not clear from reports of some experiments (e.g., Bower, 1961; MacKinnon & Russotti, 1967) whether or not there were such cues. In future work this variable should be controlled and specified.

REFERENCES

BOWER, G. H. A contrast effect in differential conditioning. *Journal of Experimental Psychology*, 1961, **62**, 196–199.

MACKINNON, F. R., & RUSSOTTI, J. S. An inverse performance pattern in a discrimination based on differential magnitudes of reward. *Psychonomic Science*, 1967, **7**, 395–396.

PERKINS, C. C. An analysis of the concept of reinforcement. *Psychological Review*, 1968, **75**, 155–172.

PERKINS, C. C. Proportion of large reward trials and within-subject reward magnitude effects. *Journal of Comparative and Physiological Psychology*, 1970, **70**, 170–172.

SEYMANN, R. G. The effect of differences in reward magnitude with correlated cues on running speed. *Journal of Experimental Psychology*, 1969, **81**, 504–509.

VALLE, F. P. Effect of exposure to feeding-related stimuli on food consumption in rats. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 773–776.

(Received April 28, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 557–560

RECOGNITION MEMORY FOR ITEM AND ORDER INFORMATION[1]

WAYNE DONALDSON[2] AND HERTA GLATHE

University of Pittsburgh

A short-term recognition memory task was designed to provide signal detection measures of the retention of both the items presented and the order in which the items occurred. The results indicate that the receiver-operating characteristic (ROC) curves for both types of information can be characterized as resulting from underlying, overlapping normal distributions. An adequate description of the retention of order information appears not to require an assumption of probabilistic incrementation based on attention to only part of that information.

The distinction between item information and order information is important in conjunction with a number of areas of investigation within the field of verbal learning and memory. It is of central concern in the work on order errors in immediate sequential recall tasks (Conrad, 1965) and to the related psychological models such as suggested by Crossman (1960). It has bearing on the research aimed at specifying the stimulus in serial learning (Young, 1968); and, in some respects, the item–order distinction is not different from the hypothesized response integration and associative learning phases of paired-associate learning (e.g., Underwood & Schultz, 1960). What is lacking is an appropriate experimental technique for separating the acquisition and/or retention of item and order information.

In a recent paper, Norman and Wickelgren (1965) developed a model which, if it can be supported, has direct implications for the distinction between item and order information. In an analysis of short-term recognition memory for single digits and pairs of digits, Norman and Wickelgren obtained fundamentally different receiver-operating characteristic (ROC) curves for single digits and digit pairs. The single-digit ROC was a smooth, continuous function, as expected, assuming the underlying strength distributions of new and old items to be Gaussian in nature. The ROC curve for digit pairs, however, appeared discontinuous, being best described by two linear components. The a posteriori probability curves actually indicated the lower limb of the ROC to be curvilinear, suggesting that *S*s were capable of ordering their confidence ratings above a low threshold. Norman and Wickelgren proceeded to develop a model based on the assumption that *S*s attend to nonoverlapping pairs of digits, the consequence being that only some of the pairs of digits are incremented in strength as a result of presentation. While this probabilistic incrementation model appears capable of handling the data, a second interpretation is possible. In constructing the digit-pair ROCs, Norman and Wickelgren plotted the conditions involving two old digits against the condition involving two new digits. These two conditions cannot be differentiated by a binary decision in that an *S* can (*a*) reject the alternative that both digits are old without accepting them both as new and (*b*) reject the possibility that both digits are new without accepting the proposition that both digits are old. The result is that the ROC plot of the condition involving two old digits against that involving two new digits compares conditions which never come directly into contact with each other in the sense that a binary decision about one of the numbers in the pair eliminates one of the digit-pair alternatives, but does not confirm the other. Further, the earlier decisions precede any decision concerning the order of two old digits. Basically, then, this comparison involves three decisions by *S*, two decisions regarding the nature of the items (item information—old or new), followed, in those cases where both digits are classified as old, by a decision regarding the order of the digits (order information—same or reverse). To measure properly the retention of order information, which is the crux of the probabilistic incrementation model, any possible confounding with item information must be removed. The present study was designed to eliminate the confounding between item and order information.

---

[1] This research was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare, to the Learning Research and Development Center, University of Pittsburgh. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions do not, therefore, necessarily represent official Office of Education position or policy.

[2] Requests for reprints should be sent to Wayne Donaldson, Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

METHOD

The method employed was similar in some respects to that used by Norman and Wickelgren (1965). On each trial, *S* saw a series of five different single digits (the target series) presented at a rate of $\frac{1}{2}$ sec/digit. This was followed after $\frac{1}{2}$ sec. by an intervening series of eight digits, sampled with replacement, and presented at the same rate. For example, *S* would see 6, 5, 1, 8, 0, followed by 6, 3, 9, 6, 4, 2, 0, 3. The digit series were filmed and presented by means of a Dunning animatic projector. Each of the 18 individually tested female undergraduates read all 13 digits out loud as they were presented. Following the intervening series, *S* turned over the top index card from the deck of cards in front of her. Printed on the card was either a single-digit or a digit-pair probe, *S* having no knowledge beforehand of which would occur. The *S* recorded the probe number(s) on her answer sheet and then recorded her response. When the probe was a single digit, *S* was required to indicate whether or not she thought that it had occurred in the target series. When the probe was a digit pair, *S* knew through the instructions that both digits had occurred adjacently in the target series and was required to indicate only whether she thought the probe pair to be in the same or reverse order as in the target series. Responses to both single digits and to digit pairs were followed by a confidence rating on a five-point scale ranging from guess to sure. The *S*s had 15 sec. in which to respond prior to the warning signal immediately preceding the presentation of the next series.

A complete replication required 18 trials, 10 trials with a single-digit probe, 8 with a digit-pair probe. Five of the single-digit probes were old items, having occurred in the target series, one from each serial position. The other five single-digit probes were new items, not having occurred in the target series. Four of the digit-pair probes were in the same order as in the target series, one from each adjacent pair of digits. The remaining four were the adjacent pairs in reversed order.

All *S*s responded to six blocks of 18 lists, each block representing a total replication of conditions. All *S*s saw the same digit lists in the same order, the order of probe cards being prepared according to 18 × 18 Latin squares such that every *S* responded under a different condition to each list. In other words, although each of the 18 *S*s saw the same list of 13 digits on Trial 1, all 18 experimental conditions were represented in the probes used for that list.

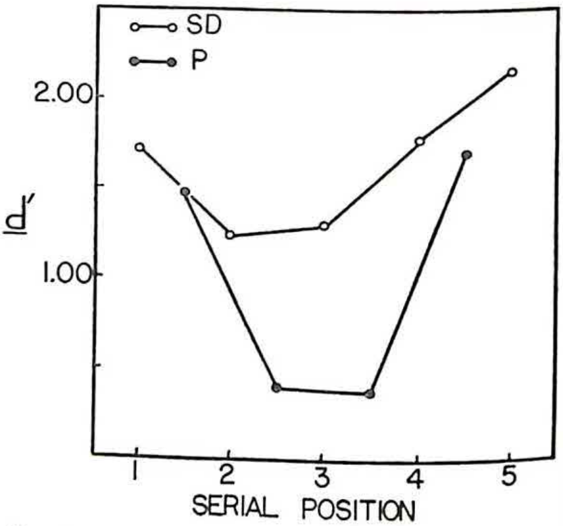


FIG. 1. *d′* for single digits (SD) and digit pairs (P) as a function of serial position.

RESULTS

Nine different ROC curves can be derived from the 18 conditions, five for the serial positions of the single digits and four for the serial positions of the digit pairs. The nine curves are all well fit by straight lines on double probability paper and all have slopes close to 1.0. Figure 1 gives the *d′* values for each of the 9 conditions as calculated from the point of intersection of the ROC with the negative diagonal. Both single digits and digit pairs show a primacy and a recency effect.

Combining over serial positions, two ROC curves, one for single digits—a cumulative plot of the conditional probability of rating *j* or stricter given old against rating *j* or stricter given new—and one for digit pairs—a plot of probability of rating *j* or stricter given same order against rating *j* or stricter given reversed order—can be obtained. These curves are shown in Fig. 2. The curve for single digits is the same smoothly changing, continuous function reported by Norman and Wickelgren (1965), although the curve in Fig. 2 appears to be slightly higher than theirs ($d' = 1.58$, as compared to $d' = 1.16$). The curve for digit pairs also appears to be a smooth, continuous function rather than the discontinuous curve reported by Norman and Wickelgren.

Figure 3 shows the a posteriori probability curves for the single digits and the digit pairs, and both appear to decrease across the full scale rather than reflect the previously found inability of *S*s to order their ratings below some threshold in the digit-pair condition.

DISCUSSION

First of all, it must be stressed that the present study does not eliminate any interaction that might exist between the retention of item and order information. It is probably true that if *S* remembers little about what items were presented to him, his performance on the order in which the items were presented would be very poor. This means that performance for order information is limited, to some extent at least, by performance on individual items. What this experiment has eliminated is the confounding between item and order information in the *measure* of order information. While the d' values for the retention of order information presented here are constrained by whatever loss of item information occurs, the measure per se is not confounded. That is, this paper provides a technique for the experimental separation of the theoretically important distinction between item and order information.

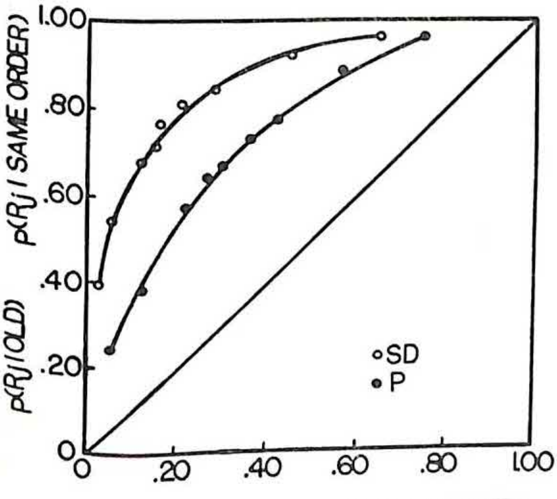


FIG. 2. ROC curves for single digits (SD) and digit pairs (P) combined over all serial positions.

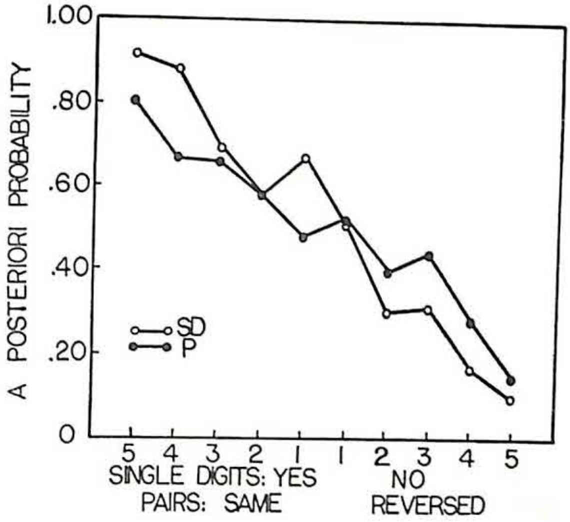


FIG. 3. A posteriori probability curves for single digits (SD) and digit pairs (P) combined over all serial positions.

Clearly, when the confounding is removed from the Norman and Wickelgren (1965) procedure, the ROC curve for retention of information regarding the order in which items occurred is continuous. This is not really surprising in the light of other data. Murdock (1965), e.g., found continuous ROCs in a recognition experiment using a paired-associate probe technique in which the X term of the incorrect A-X pairing was the B member from either the preceding or succeeding pair. In effect, this represents a task in which *S*'s decision concerns the ordering of information, to the extent, at least, that he is not concerned with whether or not an item was presented, but rather whether the items were presented in this particular order.

Finally, it is interesting to note that in a recent theoretical article by Bernbach (1967), a probabilistic model is also presented, although it is approached from the other end, viz., probabilistic forgetting, rather than incrementation, of information. Of the data in the literature, however, the ROC curves are generally well fit by a smooth, continuous function (see, e.g., the

ROCs presented by Donaldson & Murdock, 1968; and Egan, 1958), although we must heed Murdock's (1969) warning that the shape of the ROC is a very insensitive test of any model.

REFERENCES

Bernbach, H. A. Decision processes in memory. *Psychological Review,* 1967, **74**, 462–480.

Conrad, R. Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 161–169.

Crossman, E. R. F. W. Information and serial order in human immediate memory. In C. Cherry (Ed.), *Information theory.* London: Butterworth, 1960.

Donaldson, W., & Murdock, B. B., Jr. Criterion change in continuous recognition memory. *Journal of Experimental Psychology,* 1968, **76**, 325–330.

Egan, J. P. Recognition memory and the operating characteristic. (Tech. Rep. No. AFCRC-TN-58-51) Bloomington, Ind.: Indiana University, Hearing and Communication Laboratory, 1958.

Murdock, B. B., Jr. Signal-detection theory and short-term memory. *Journal of Experimental Psychology,* 1965, **70**, 443–447.

Murdock, B. B., Jr. Short-term memory for associations. In D. A. Norman (Ed.), *Models of memory.* Englewood Cliffs, N. J.: Prentice-Hall, 1969.

Norman, D. A., & Wickelgren, W. A. Short-term recognition memory for single digits and pairs of digits. *Journal of Experimental Psychology,* 1965, **70**, 479–489.

Underwood, B. J., & Schultz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

Young, R. K. Serial learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

(Received April 29, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 561–565

RESPONSE LATENCY AS A FUNCTION OF INTERSTIMULUS INTERVAL IN CONDITIONED EYELID DISCRIMINATION[1]

WILLIAM E. VANDAMENT[2]

State University of New York at Binghamton

Two laterally separated CSs, one with a 400-msec. interstimulus interval (ISI) and the other with an 800-msec. ISI, were presented in a random sequence to each experimental *S* in an eyelid conditioning experiment. The majority of experimental *S*s gave more short-latency CRs to the 400-msec.-ISI CS than to the 800-msec.-ISI CS and more long-latency than short-latency CRs to the long-ISI CS. The distributions of CR latencies appeared fairly specific to each interval after 80 trials, with some indication that the CS with the 800-msec. ISI elicited both conditioned and generalized responses.

Latency of the eyelid CR has been shown to covary with the interstimulus interval (ISI) in several studies (Boneau, 1958; Ebel & Prokasy, 1963; Prokasy, Ebel, & Thompson, 1963). In the study by Prokasy et al., latencies increased monotonically with progressive shifts to longer CS-UCS intervals throughout the 630–2,497 msec. range investigated. Ebel and Prokasy, using intervals of 200, 500, and 800 msec., observed appropriate latency changes whether shifts were made in the direction of increasing or decreasing the ISI. In these investigations, each ISI was held constant for relatively long segments of acquisition, i.e., for 40–60 trials in the Prokasy et al. study and for blocks of 400 trials in the Ebel and Prokasy study.

Kimble, Leonard, and Perlmuter (1968) analyzed latency characteristics of CRs in an investigation of the ISI and between- vs. within-*S*s designs. The random alternation of one CS having an ISI of 500 msec. with a second CS at a 1,200-msec. ISI in the within-*S*s condition was found to increase the number of short-latency CRs given to the long-ISI stimulus over that observed when intervals were manipulated between *S*s. Further, the frequency of such short-latency CRs diminished rapidly when the CS with the short ISI was no longer reinforced, leading to the conclusion that these responses were generalized from the short-ISI stimulus.

These results suggest that if two highly discriminable CSs, having different CS-UCS intervals, were presented in a random sequence, as in the typical differential conditioning situation, *S*s might learn to respond with CRs of a characteristic latency to each CS. In the present investigation, latency and frequency measures were obtained for two such laterally separated CSs, one having a 400-msec. ISI and the other an 800-msec. ISI. A control group was given training with one CS having a 400-msec. ISI and the other CS, a 2,500-msec. ISI found not to produce CRs during the initial 800-msec. period (Prokasy et al., 1963). This control condition provides an 800-msec. scoring interval for the 2,500-msec.-ISI CS which should contain generalized responses and spontaneous blinks as a base line for evaluating responses to the 800-msec.-ISI CS of the experimental condition.

METHOD

Apparatus.—The *S*s were seated in a straight-back chair with cushioned chinrest in a sound-treated, semidarkened room facing an enclosed 30 × 30 in. flat black panel on which the visual stimuli, increases in luminance of one of two milk-glass disks, were presented. These disks of 5-cm. diameter were separated laterally at their centers by 15 cm. and were centered on the panel, which was approximately 40 in. from the *S*'s eyes. The enclosure was 46 in. in depth and was also flat black. Ambient illumination and framing of the stimulus panel was provided by two clear Plexiglas

---

[1] This study was supported in part by National Science Foundation Grant GB-4492. Data analyses were performed by the Computing Center, State University of New York at Binghamton.

[2] Requests for reprints should be sent to William E. Vandament, Department of Psychology, State University of New York at Binghamton, Binghamton, New York 13901.

rods of .25-in. diameter which extended 25 in. in a horizontal plane at the top and bottom of the panel. These rods were each lighted by two 7-w. bulbs placed at their ends and were of .02–.03 apparent ftc. along their surfaces. A visual ready signal and fixation point consisted of the onset of a .5-in. diameter green indicator light of .95 apparent ftc. located midway between the visual CSs. A white-noise generator was used to mask apparatus sounds from the adjacent *E*'s room. The UCS, a corneal air puff, was delivered via a compressed air-tank–solenoid-valve system, and eyelid movements were recorded by means of a potentiometer-polygraph system described elsewhere (Vandament, 1969). Duration of stimuli and ISIs were controlled by a series of Hunter decade interval timers.

Procedure.—For experimental *S*s, one CS was presented with a CS-UCS interval of 400 msec. (short-ISI CS) and the second CS, with an 800-msec. interval (long-ISI CS). In the control condition, corresponding intervals were 400 and 2,500 msec. In both conditions, the CSs consisted of increases in luminance of one of the milk-glass disks from .06 to 1.60 apparent ftc. and were of 400-msec. (short-ISI CS) and 800-msec. (long-ISI CS) duration. The UCS was a 2-psi air puff of 50-msec. duration delivered to the right eye.

A series of 50 trials to each of the two CSs was presented using a random sequence within the restrictions of equal occurrence in blocks of 10 trials and of each CS having been preceded equally often by a short- or long-ISI CS in blocks of 20 trials. Intertrial intervals varied nonsystematically from 15 to 25 sec. about a mean of 20 sec., with the ready-signal–CS intervals having a mean of 3 sec. and a 2–4 sec. range.

Subjects and experimental design.—The *S*s were 24 male and 16 female students from an introductory psychology course at Harpur College. Half of the *S*s were assigned to each of the two basic conditions. Counterbalancing the spatial position of the short-ISI CS and placement of the short-ISI CS in the sequence left 5 *S*s assigned nonsystematically to each of eight cells within a restriction of a 3:2 male-to-female ratio.

Definition of response. —Deflections of 1 mm. or more of the recording pen corresponding to approximately .5-mm. lid movement within a latency period of 160 msec. following CS onset to CS offset were considered CRs. Response frequency data were recorded for the short-ISI CSs and separately for the initial and final 400-msec. segments of the long-ISI CSs. Since more than one CR per trial was recorded if the pen was deflected a second time after having returned to the base line for at least 40 msec. following the initial CR, it was possible for CRs to be recorded during both 400-msec. periods on each 800-msec. trial. This scoring method was employed on the a priori grounds that since more than one ISI was used: (*a*) recording only the first response to the long-ISI CS could conceivably mask the latency of a CR to that CS if a generalized response from the short-ISI CS had occurred, and (*b*) recording only those responses which overlapped the UCS (cf. Prokasy et al., 1963) might eliminate generalized responses from the short-ISI CS and result in spuriously efficient latency differentiation.

RESULTS AND DISCUSSION

Percentages of CRs given to the CSs with the short ISI and to the two segments of the long-ISI stimuli in blocks of 20 trials appear in Fig. 1. These data, as well as the latency data which appear later, contain responses from four experimental and three control *S*s who met a slope criterion for "voluntary" responding. Although CR latencies for these *S*s were on the average lower than the latencies of other *S*s, no gross deviations in relative latency or frequency characteristics to the two stimuli were noted.

Two indexes of differentiation are possible in this situation: (*a*) a comparison of the CSs with regard to responses occurring prior to 400 msec. and (*b*) the relative numbers of short- and long-latency responses given to the 800-msec.-ISI stimulus. These are subsequently referred to as short-latency and long-ISI indexes, respectively.

An analysis of variance related to short-latency differentiation was performed in which the single between-*S*s variable was acquisition group (experimental vs. control), while within-*S*s effects consisted of CS-UCS interval (short-ISI CS vs. 160–400 msec. of the long-ISI CS) and trial blocks. A significant ISI effect, $F\ (1, 38) = 37.84$, $p < .001$, verifies the reliability of the higher response levels which are observable for the short-ISI stimuli in each acquisition group. The absence of a significant Groups × ISI interaction and the finding that 19 of the 20 experimental *S*s and 18 of the 20 control *S*s gave more CRs to the short-ISI CS than to the initial portion of the long-ISI CS suggest a short-latency differentiation was present for both groups.

The initial segments of the long-ISI CSs display increments in responding only through the third block of trials. The short-ISI stimuli, however, are characterized by increases in response levels to the close of acquisition leading to a significant ISI × Trials interaction, $F\ (4, 152) = 8.68$, $p < .001$. The increments observed from the fourth to the fifth blocks in the short-ISI functions suggest that learning may not have been completed in the 100 trials given so that conclusions cannot be drawn regarding asymptotic response levels.

As can be observed in Fig. 1, response levels during the 401–800 msec. portion of the long-ISI stimulus increase with trials for the experimental group in the monotonic fashion usually associated with an acquisition process. An analysis of variance was performed on numbers of responses given during the final 400 msec. of the long-ISI stimulus in the experimental and control conditions, with a resulting groups effect, F (1, 38) = 28.69, $p < .001$, and Groups × Trials interaction, F (4, 152) = 9.11, $p < .001$. The increments observed during this 401–800 msec. period in the experimental condition were also associated with a preponderance of long-latency responding to the 800-msec.-ISI stimulus, with 17 of the 20 experimental *S*s giving more than 50% of their responses to that stimulus during the second 400 msec. An analysis of variance comparing response levels during the 160–400 and 401–800 msec. portions of the long-ISI CS for experimental *S*s was significant, F (1, 19) = 4.48, $p < .05$, providing further evidence that a long-latency differentiation was present.

Another view of what appears to be an acquisition process for long-latency CRs in the experimental condition is given by Fig. 2, in which latency distributions to the long-ISI CSs are plotted for two periods of acquisition. The lack of congruence between the experimental and control conditions during the final 400 msec. of Trials 81–100 is related to the development of a distinct 481–560 msec. modal range in the experimental group which is coincident with increased long-latency responding in that condition.

The similarity of the experimental and control latency distributions for 1–400 msec. during Trials 81–100 in Fig. 2 suggests further that some portion of the short-latency CRs in the experimental group can be related to generalization from the short-ISI CS. Additional information in this regard

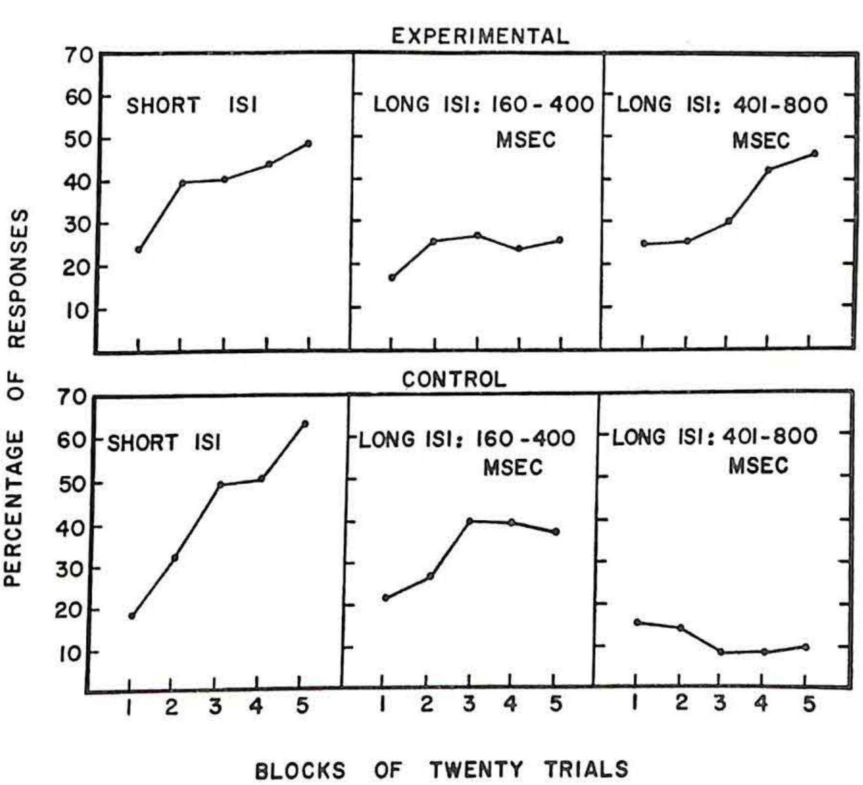


FIG. 1. Percentages of CRs to the short-ISI CS and to the long-ISI CS for experimental and control groups in five blocks of 20 acquisition trials (10 to each CS).

is based on findings in standard differential conditioning experiments that response probability to the CS^+ and CS^- for a given, or *n*th, trial is related to reinforcements or nonreinforcements on Trial $n-1$ (Moore, 1964; Prokasy & Allen, 1969; Vandament, 1969). In the present experimental condition for Trials 21–100, short-latency CR percentages to the 800-msec.-ISI CS following 400- and 800-msec. trials were 29% and 20%, respectively, with long-latency percentages of 32% and 40% observed. Analysis of variance on these data yielded a Latency of Trial $n \times$ CS of Trial $n-1$ interaction, $F(1, 19) = 4.50$, $p < .05$. This result is consistent with the Kimble et al. (1968) finding that the frequency of short-latency responses to a long-ISI CS covaried with the presence or absence of reinforcement to a short-ISI CS which was alternated with the long-ISI stimulus. Thus, although other factors may also have been involved in the production of short-latency CRs to the long-ISI CS, existing evidence indicates that generalization from the short-ISI CS is responsible for at least a portion of such responding.

Individual response patterns for experimental *S*s show positive relationships between all three scoring intervals when the data of 3 deviant *S*s (who can be classified as "short-latency" *S*s by a 50% criterion applied to the long-ISI stimulus data) are disregarded. For the remaining 17 *S*s, a Pearson r of .70, $p < .01$, was obtained between response levels to the short-ISI CS and to 160–400 msec. of the long-ISI CS; an r of .91, $p < .001$, between short-ISI CS and 401–800 msec. responding; and an r of .75, $p < .001$, between response levels in the two segments of the CS with the 800-msec. ISI. These results indicate a factor reflecting "responsiveness" to be present in the data at all scoring intervals. Further, while some *S*s were characterized by a short-latency response pattern to both stimuli, corresponding long-latency patterns involving low response levels to the short-ISI CS and high levels to the final 400 msec. of the long-ISI CS were not observed. In fact, ignoring the data of the 3 short-latency *S*s, a

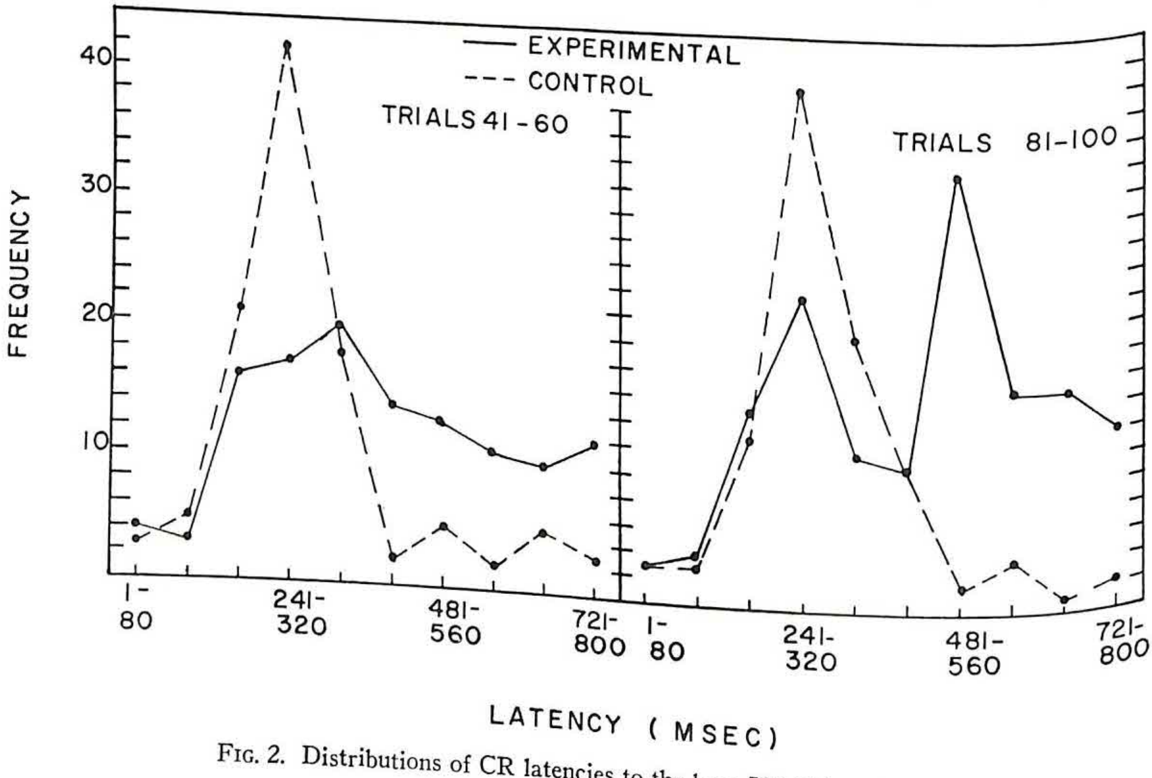


FIG. 2. Distributions of CR latencies to the long-ISI CSs during two periods of acquisition.

positive relationship between short-ISI responding and long-ISI differentiation (percentage of long-latency minus percentage of short-latency CRs to the long-ISI CS) was found, $r = .54$, $p < .02$.

The short-latency and long-ISI differentiations observed in the present study differ somewhat from results obtained by Kimble et al. (1968) for their within-*S*s manipulation of the CS-UCS interval. This condition of Kimble et al., in which CS-UCS intervals of 500 and 1,200 msec. were used, did not result in a short-latency differentiation, as in the present study. Although not analyzed for long-ISI differentiation, the Kimble et al. data do suggest a dominance of long-latency over short-latency responding to the long-ISI stimulus, as found in the present experiment.

An additional difference between the two studies involves the relationship of short-latency responses to long-latency responses for a given long-ISI trial. In the present study, the occurrence of either a long- or short-latency response tended to preclude the occurrence of the other on the same trial. Two responses were scored on the same trial for 4% and 1% of the long-ISI trials in the experimental and control conditions, respectively. During Trials 81–100, only 17.6% of the short-latency responses to the CS with the 800-msec. ISI were followed by a long-latency CR in the experimental condition, while the long-latency response rate was 56% in the absence of a short-latency response. In the Kimble et al. (1968) investigation, 32% of the long-latency CRs to the stimulus with the long ISI followed a short-latency response given on the same trial. It is possible that the 400-msec. difference between the intervals in the present study did not allow sufficient time for two responses to occur with high probability, given the scoring criterion which required a return to base line following the initial CR for the second response.

In summary, the present investigation has demonstrated that it is possible for *S*s to respond differentially to CSs whose reinforcement properties differ only with regard to CS-UCS interval. Further investigation is indicated to determine if such results are specific to the procedures employed in the present study and whether short- and long-latency responses are necessarily either positively or negatively related within a given trial, as suggested by existing data.

REFERENCES

BONEAU, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *Journal of Experimental Psychology*, 1958, **56**, 464–471.

EBEL, H. C., & PROKASY, W. F. Classical eyelid conditioning as a function of sustained and shifted interstimulus intervals. *Journal of Experimental Psychology*, 1963, **65**, 52–58.

KIMBLE, G. A., LEONARD, T. B., III, & PERLMUTER, L. C. Effects of interstimulus interval and discrimination learning in eyelid conditioning using between- and within-*S*s designs. *Journal of Experimental Psychology*, 1968, **77**, 652–660.

MOORE, J. W. Differential eyelid conditioning as a function of the frequency and intensity of auditory CSs. *Journal of Experimental Psychology*, 1964, **68**, 250–259.

PROKASY, W. F., & ALLEN, C. K. Instructional sets in human differential eyelid conditioning. *Journal of Experimental Psychology*, 1969, **80**, 271–278.

PROKASY, W. F., EBEL, H. C., & THOMPSON, D. D. Response shaping at long interstimulus intervals in classical eyelid conditioning. *Journal of Experimental Psychology*, 1963, **66**, 138–141.

VANDAMENT, W. E. Differential eyelid conditioning as a function of CS-UCS interval and distance separating the CSs. *Psychological Reports*, 1969, **25**, 407–411.

(Received April 29, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 566–570

REACTION TIME AS AN INDEX OF REHEARSAL IN SHORT-TERM MEMORY[1]

ROBERT F. STANNERS,[2] GARY F. MEUNIER, AND DONALD B. HEADLEY
Oklahoma State University

A short-term memory (STM) task was used in which the materials were visually presented trigram triads of two types, either easy (E-Pr) or difficult (D-Pr) to pronounce. A pretesting procedure was employed which closely equated the performance of all *S*s at a 0-sec. retention interval. Following the pretesting, experimental (E) groups received trials consisting of the presentation of the triad, a rehearsal period of 7 sec., and a recall signal. At one of four different times (1, 2, 4, or 6 sec.) after the offset of the triad, a buzzer was sounded to which *S* was to respond as quickly as possible by moving a toggle switch. Control (C) conditions involved only the RT task and not the STM task. E groups displayed substantially longer RTs than did the C groups at all levels of buzzer interval (BI). Average RT for the E groups decreased systematically over the levels of BI, while the trend lines for the C groups were essentially horizontal. The E group with the D-Pr material displayed significantly longer RTs than did the E group with the E-Pr material, while C groups were essentially the same. The results were interpreted as supporting the existence of a rehearsal process having the following properties: Rehearsal load decreases over time following presentation of the materials and is lower for E-Pr than for D-Pr materials.

Rehearsal is a process assumed to take place over a period of several seconds following the presentation of material in a learning or memory task. The vulnerability of short-term memory (STM) performance to various types of activities which are interpolated between the presentation of the material and recall has been the major operational basis for a rehearsal process. A task introduced by Peterson and Peterson (1959) convincingly demonstrated that substantial amounts of short-term forgetting could be produced by using backward counting as an interpolated activity. Since then, a very large number of experiments have used backward counting or similar types of activities as a means of rehearsal prevention. Another way of inferring rehearsal from recall performance has been suggested in a recent experiment by Stanners and Meunier (1969), in which the initial acquisition of the materials was held to about 68% and rehearsal intervals were inserted prior to the interpolated activity. The overall gain in recall performance stemming from a 5-sec. period of rehearsal as compared to a 0-sec. period was about 64% in the case of easy-to-pronounce trigram triads. It appears that when initial acquisition is well below 100%, a short rehearsal period can exert a very substantial effect on recall.

Crowder (1967) has pointed out that the type of STM task exemplified by the Peterson and Peterson (1959) study bears important similarities to methods used in assessing division of attention. The basis for this assertion is that during the retention intervals of STM tasks, there is presumed to be competition between rehearsal and the interpolated task. The Crowder experiment involved recall for five-word stimuli after 24-sec. intervals filled with a self-paced key-pressing task. A rehearsal process was inferred from the fact that the number of key presses was lower on the memory trials than on control trials, for which no recall task was involved.

Another approach to the assessment of rehearsal was suggested in a study by Guilford and Ewart (1940). Advertisements from popular magazines were shown to *S*s with the instructions that they would later be tested on the material. At varying times (3–6, 12–15, or 21–25 sec.) after the onset of the picture, an auditory stimulus was presented, to which *S* was supposed to react as quickly as possible with a key release. Reaction times (RTs) were systematically longer while *S* was studying the picture as compared to a control condition, which was a simple RT to the same auditory stimulus. It was also found that RT while studying the picture decreased from the 3–6 sec. period to the 12–15 sec. period.

---

[1] This study was principally supported by a grant from the Oklahoma State University Research Foundation.

[2] Requests for reprints should be sent to Robert F. Stanners, Department of Psychology, Oklahoma State University, Stillwater, Oklahoma 74074.

A technique similar to that of Guilford and Ewart (1940) was employed in the present experiment to study rehearsal in a STM task. Interest was focused on rehearsal intervals of 1–6 sec. because a previous experiment (Stanners & Meunier, 1969) indicated that the major performance changes occurred within a 5-sec. rehearsal period. An additional question was whether or not the RT measure would reveal differences in rehearsal for materials of different levels on the scale of pronounciability constructed by Underwood and Schultz (1960).

METHOD

Subjects.—The *S*s were 112 right-handed undergraduates at Oklahoma State University. The students received a small credit toward their final course grades as an inducement to participate.

Materials.—Easy-to-pronounce (E-Pr) and difficut-to-pronounce (D-Pr) sets of trigrams were selected from the pool scaled by Underwood and Schulz (1960). Within each set, the trigrams were organized into 38 triads so that the letter similarity within triads was very low and the range of average Pr values within each set was minimal. A detailed account of the selection and organization of the trigrams may be found in Stanners and Meunier (1969). The triads were typed horizontally on Radio Mats with a single space between each trigram and mounted in 35-mm. slide holders.

Procedure and design.—The *S* was seated at a distance of 2 ft. from a 12 × 12 in. Plexiglas screen onto which the triads were back-projected to a height of 1.5 in. The basic presentation equipment consisted of two Kodak Carousel projectors equipped with solenoid-operated shutters and programmed by an eight-bank Lafayette timer. A Hunter Klockcounter was used to measure the RTs.

Underwood (1964) has pointed out that retention studies comparing different types of material must employ acquisition levels below 100% to avoid the possibility of differential overlearning effects. The purpose of the first (pretesting) portion of the experiment was to establish an exposure duration that would place each *S* at a standard level of performance (six to eight correct letters in correct positions) for a 0-sec. retention interval. The materials used in pretesting were 10 slides which were representative of the Pr values within a particular Pr group. The exposure time for the first slide of the pretesting series was determined from pilot data and was .5 sec. for the E-Pr material and 5.0 sec. for the D-Pr triads. The screen was illuminated with a homogeneous blank field both preceding and following the presentation of a slide. A red pilot lamp mounted below the screen was used to signal the beginning of the recall period and came on with the offset of the triad. The *S* was instructed to write his recall on a scoring pad with spaces corresponding to the spatial arrangement of the slide. Over the nine trials of pretesting which followed the initial slide, the exposure interval was increased or decreased until *S* reached the performance criterion of six to eight correct letters in correct positions. A 10-sec. maximum for exposure times was established, and two *S*s in the D-Pr group were replaced on this basis. The *S*s were allowed 10 sec. to write their recall, which was followed by a 20-sec. intertrial interval. Each new trial was preceded by a warning signal constituted by two blinks of a white pilot lamp with a 1-sec. interval between blinks. All *S*s received 10 pretesting trials, regardless of the number of trials (fewer than 10) actually required to establish the criterion time. The exposure time which produced the criterion performance was used throughout the second portion of the experimental session.

In the second phase of the experiment, the E-Pr and D-Pr experimental groups (E-PrE and D-PrE) received the following procedure. A triad was shown for the duration established during pretesting. The offset of the slide was followed by a 7-sec. (rehearsal) period before the onset of the recall signal. At one of four different times (1, 2, 4, or 6 sec.) during the rehearsal interval, a buzzer was sounded to which *S* was to respond as quickly as possible by moving a toggle switch held between thumb and forefinger of the right hand. The buzzer intervals (BI) were scheduled randomly, with some constraints, to be noted later, in the comments on the balancing of the experimental variables. The only instructions *S*s were given concerning the rehearsal period were that their main task was to remember the material accurately and that they were to respond as quickly as possible when the buzzer sounded. The combined recall period and intertrial interval was 30 sec.

A finding by Foley (1959) that RT can be affected by the onset of the foreperiod indicated that two control groups (E-PrC and D-PrC) should be used. If the presentation of the triads for the E groups is considered the foreperiod signal, there would be a substantial difference between the E-Pr and D-Pr groups in the duration of the signal. Each experimental (E) *S* had a control (C) counterpart. A given C *S* received a foreperiod signal

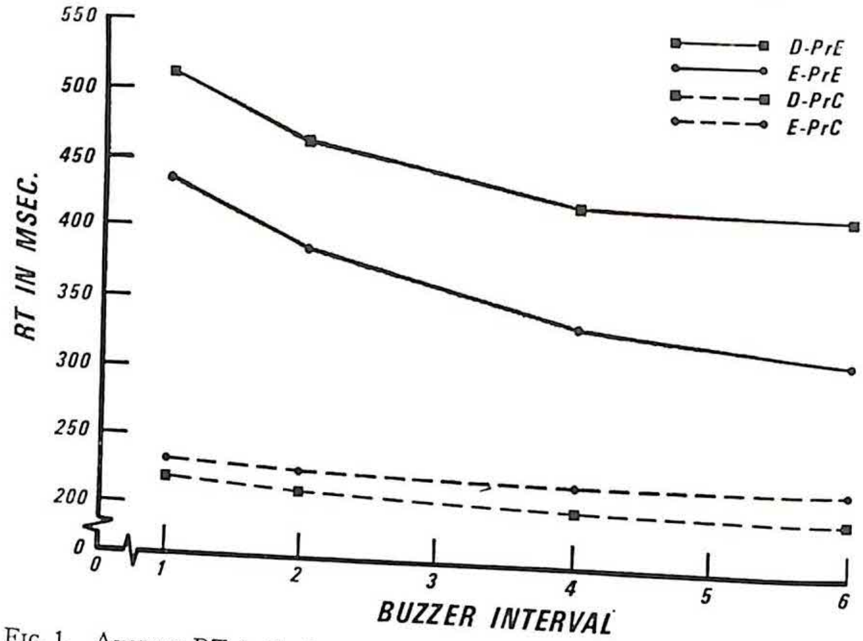


FIG. 1. Average RT to the buzzer as a function of group and buzzer interval.

(an amber light flashed on the screen) of the same duration as the triad duration of the matching E S. The scheduling of the buzzer intervals was the same for an E-C pair of Ss. The C Ss received the same pretesting procedure as the E Ss. At the beginning of the second part of the experimental session, C Ss were told that there would be no further testing of memory and that their only task was to respond as quickly as possible to the buzzer.

The two levels of Pr and the E and C conditions were orthogonal, between-Ss conditions, with 28 Ss used in each of the four subgroups. The Ss were randomly assigned to the subgroups, with the restriction that the Nth S not be assigned to a given subgroup until $N-1$ Ss had been assigned to all other subgroups. There were 14 Ss of each sex within each of the four subgroups. Buzzer interval was a within-Ss variable, with seven slides assigned to each of the four levels. The major features of the balancing procedure were as follows: Each slide occurred equally often at each of the 28 serial positions and at each level of BI. Twenty-eight different orders of the BIs were used.

RESULTS

Pretesting data.—The E-PrE and D-PrE groups were compared on initial acquisition level. The scores were the number of correct letters in correct positions on the last three trials of pretesting. The groups were highly similar in mean performance, 6.70 correct letters for the D-PrE group and 6.80 for the E-PrE group. The .05 level was adopted as the minimum level of significance for all statistical tests; using this standard, a comparison of the pretesting means by t test fell far short of significance.

The mean exposure times required to produce the criterion level of recall performance were .55 sec. for the E-PrE group and 6.04 for the D-PrE group. No statistical comparison was necessary since the highest time in the E-PrE group was .90 sec. and the lowest time in the D-PrE group was 2.00 sec.

RT data.—Four RT scores, which were the mean RTs taken over seven trials at each level of BI, were calculated for each S.[3] The overall analysis of these data was a $2 \times 2 \times 4$ analysis of variance with Pr and E-C as between-Ss variables and BI as a within-Ss variable. The significant main effects were Pr, F (1, 108) = 4.03, $p < .05$; E-C, F (1, 108) = 118.85, $p < .001$; and BI, F (3, 324) = 39.57, $p < .001$. The means for all combinations of conditions are depicted in Fig. 1 and suggest two interaction effects. The mean RT for the D-PrE condition is markedly slower than for the E-PrE condition, while the means of the E-PrC and D-PrC groups are highly similar. The BI subgroup means in the E conditions display a very definite downward trend, while in the C groups, the trend lines are virtually horizontal. Both interactions were detected in the analysis of variance; Pr × E-C, F (1, 108) = 8.38, $p < .005$; and E-C × BI, F (3, 324) = 23.31, $p < .001$.

---

[3] The median RTs for each S were also calculated. The pattern of the subgroup means of medians was highly similar to that for the means of means, which is depicted in Fig. 1. The statistical analyses for the medians produced results identical to those for the means.

Answers to the initial questions concerning the role of duration of the rehearsal interval and the pronunciability of the material required subanalyses. A two-factor analysis of variance was performance at each level of the E-C variable. In the E condition, there was a significant effect of Pr, F (1, 54) = 6.62, $p < .025$; and BI, F (3, 162) = 36.37, $p < .001$; but no significant interaction effect. In the C condition, the only significant effect was that of BI, which was very small in absolute value, but nevertheless exceeded the .025 level, F (3, 162) = 3.57.

Recall data.—The descriptive statistics for the recall data are presented in Table 1. It should be kept in mind that all recall measures were taken after a 7-sec. rehearsal period; the conditions were differentiated only by the intervention of the buzzer. A two-factor analysis of variance performed on the recall scores indicated only a significant effect of Pr, F (1, 54) = 6.39, $p < .025$. The fact that the E-PrE group was superior in recall performance to the D-PrE group after the rehearsal period but not during the pretesting indicated that a performance change might have occurred during the rehearsal period for one or both of the groups. The nature of this change was investigated by comparing the mean recall over all four levels of BI with the mean of the last three pretesting trials. The pretesting mean for the E-PrE group was 6.80, while the overall mean after the 7-sec. rehearsal period was 7.29. A t test for correlated measures produced a value of 2.71, $p < .025$. The comparable means for the D-PrE group were 6.70 and 6.71; the t test was far short of a significant value.

DISCUSSION

The RT measure was sensitive to several aspects of the retention task. The slowdown in RT produced by the memory task was very substantial; there was little overlap in the distributions of the individual S scores of the E and C groups. This result provides independent support for the performance data, which indicates a rehearsal process. Other findings in the present experiment provide information on specific properties of the rehearsal process. In a previous experiment (Stanners & Meunier, 1969), recall data indicated that the rehearsal process was more efficient with E-Pr material than it was with D-Pr material. The separation of the E-PrE and D-PrE groups in the present study leads to the same conclusion. A theory of STM advanced by Waugh and Norman (1965) proposed that one of the functions of rehearsal was to transfer material from a memory state of very limited capacity (primary memory) to a much more commodious state (secondary memory). The substantial decreases in RT over the levels of BI for the E groups support the Waugh and Norman conception by suggesting that S's rehearsal load decreased over successive values of the rehearsal interval. However, no differential decrease in RT occurred for the E-PrE group, which would have been expected on the assumption that the transfer of E-Pr material from primary memory to secondary memory would be more efficient than the transfer process for D-Pr material. A possible reason for the failure to find divergence of the RT functions is that RT was not tested soon enough in the retention interval. The first value of BI was 1 sec., by which time enough E-Pr material may have been transferred to produce a substantial reduction in RT. Although a 1-sec. rehearsal interval appears to be very short, it is almost twice as long as the average exposure duration (.55 sec.) for the E-Pr material.

The analysis of the recall data indicated that recall performance for the E-PrE group was better at a 7-sec. retention interval than it was during pretesting at a 0-sec. interval. The same effect did not appear for the D-PrE group. These findings cannot be given a com-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE RETENTION SCORES

Statistic	Buzzer interval			
	1 sec.	2 sec.	4 sec.	6 sec.
E-PrE				
M	7.05	7.26	7.45	7.43
SD	.79	.99	.92	.88
D-PrE				
M	6.63	6.73	6.75	6.74
SD	1.14	1.12	.92	1.33

pletely unambiguous interpretation as a reminiscence effect, because different materials were used in pretesting than were used in the experiment proper. However, the results are highly similar to those of a previous experiment (Stanners & Meunier, 1969) in which virtually the same set of materials was used and in which a reminiscence effect was found for the E-Pr but not for the D-Pr material.

The RT measure seems to offer some promise as a means of evaluating inferences about rehearsal drawn from recall data. Other implicit processes such as mediation may also be amenable to study by this technique.

REFERENCES

Crowder, R. G. Short-term memory for words with a perceptual-motor interpolated activity. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 753–761.

Foley, P. J. The foreperiod and simple reaction time. *Canadian Journal of Psychology,* 1959, **13**(1), 20–22.

Guilford, J. P., & Ewart, E. Reaction-time during distraction as an indicator of attention-value. *American Journal of Psychology,* 1940, **53**, 554–563.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology,* 1959, **58**, 193–198.

Stanners, R. F., & Meunier, G. F. Pronounciability and rehearsal time in short-term memory with controlled acquisition. *Journal of Experimental Psychology,* 1969, **80**, 359–363.

Underwood, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3**, 112–129.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review,* 1965, **72**, 89–104.

(Received May 2, 1969)

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Invariance of Total Learning Time under Different Conditions of Practice: Rose T. Zacks*: Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122.

Whole-Part Transfer from Paired-Associate to Free Recall Learning: Gordon Wood*: Department of Psychology, Olds Hall, Michigan State University, East Lansing, Michigan 48823.

Visual and Auditory Intrusion Errors in Short-Term Memory: Kenneth R. Laughery* and Gilbert J. Harris: Department of Industrial Engineering, Parker Engineering, State University of New York at Buffalo, Buffalo, New York 14214.

Effects of Contextual Similarity on Unlearning in the A-B, D, E, F and B, D, E, F Paradigms: Richard K. Landon and James H. Crouse*: College of Education, University of Delaware, Newark, Delaware 19711.

Proactive Inhibition in Short-Term Memory: Jean E. Poppei, Barbara L. Finlay, and W. H. Tedford, Jr.*: Department of Psychology, Oberlin College, Oberlin, Ohio 44074.

Forgetting: Interference or Decay?: Dominic W. Massaro*: Department of Psychology, University of California, San Diego, P.O. Box 109, La Jolla, California 92037.

Clustering in Free Recall Following Paired-Associate Learning: Peggy A. Runquist*: Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

Optimal Potentiating Effects and Forgetting-Prevention Effects of Tests in Paired-Associate Learning: Chizuko Izawa*: Department of Psychology, 107 Townsend Hall, State University of New York at Buffalo, Buffalo, New York 14214.

Verbal Interference with Encoding in a Perceptual Classification Task: Howard S. Hock and Howard Egeth*: Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218.

Discrimination Influences of the Postchoice Display of Incorrect Objects: F. Robert Treichler* and Brenda B. Riccio: Department of Psychology, Kent State University, Kent, Ohio 44240.

Prevention of Unlearning in Verbal Transfer by First-List Reinstatement: John Jung*: Department of Psychology, California State College at Long Beach, Long Beach, California 90801.

Effects of Repetition and Exposure Duration on Memory: Douglas L. Hintzman*: Department of Psychology, University of Oregon, Eugene, Oregon 97403.

Accuracy and Latency in Short-Term Memory: Evidence for a Duel Retrieval Process: Stanley M. Moss* and Jo-Anne Sharac: Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Reversal Learning in a Successive Discrimination Using Intermittent Reinforcement: Roger L. Mellgren* and John W. P. Ost: Department of Psychology, Psychology Building, Indiana University, Bloomington, Indiana 47401.

---

* Asterisk indicates author for whom address is given.

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 571–572

SUPPLEMENTARY REPORTS

RECEPTION VERSUS SELECTION PROCEDURES IN CONCEPT LEARNING

FRANK S. MURRAY[1] AND ROBERT E. GREGG
Randolph-Macon Woman's College *Eastern Kentucky University*

Performance efficiencies of two methods of presenting concepts for learning were compared. Adult *S*s learned simple concepts by dividing wooden blocks into groups according to shape or color. The selection procedure consisted of presenting simultaneously the total array of stimulus patterns, and *S* selected instances of the concepts from this array. The order of selecting instances by *S*s in the selection group was presented successively and individually to yoked *S*s in the reception group. The results showed that concept learning was more efficient with the reception procedure.

One of two different procedures of presenting instances of concepts is usually used in studies of concept learning: a reception procedure or a selection procedure. With the reception procedure, instances of concepts are usually presented successively and individually, and *S* is required to indicate whether the instance is, or is not, the concept. With the selection procedure, all instances are usually presented simultaneously, and *S* is allowed to choose the order of selecting instances of the concept (Bruner, Goodnow, & Austin, 1956). Few studies have attempted to compare the efficiency of both procedures.

An attempt to compare the two procedures was performed by Huttenlocher (1962). She used seventh-grade children as *S*s and a selection procedure in which the array of possible stimulus patterns was not available to *S*s for continuous inspection. Her results showed that *S*s in a nonmanipulating (reception) group were superior to those in a manipulating (selection) group.

In a more recent study, Hunt (1965), using adult *S*s (college students), reported results in which the selection procedure was more effective than the reception procedure. In this study, *S*s were required to use grammatical rules and conjunctive concepts to generate instances of concepts.

Besides *S*s, other differences existed between the two studies. Huttenlocher (1962) required her *S*s to manipulate only two or three blocks to form instances of concepts, and in her selection procedure, the array of possible stimulus patterns was not available to *S*s for continuous inspection. Hunt's (1965) experiment was performed in three parts: preliminary, learning, and test phases. Only errors made by *S*s in the test phase were counted. In the Huttenlocher study, errors made in a learning stage (the only phase) were counted.

The purpose of the present experiment was to compare the performance efficiency of the two procedures using adult *S*s and a selection procedure in which the array of possible stimulus patterns was available for continuous inspection. Instances of concepts selected by *S*s with the selection procedure were presented successively and individually to yoked *S*s in the reception group. As in the Huttenlocher (1962) and Hunt (1965) studies, *S*s in the selection procedure were required to select stimulus patterns and make commitments concerning their categories. A simple concept learning task was used to evaluate the two procedures.

Method.—Forty students (17 males and 23 females) enrolled in introductory psychology courses at Eastern Kentucky University volunteered to participate as *S*s.

Sixteen wooden blocks, varying in shape and color, were used. Four of the blocks were circles, four were squares, four were triangles, and four were crosses. There were four colors of each shape: red, green, blue, and yellow. The blocks were 3.81 cm. in length and width and 1.27 cm. in depth.

A yoked design was employed. Of the 40 *S*s, 20 were randomly assigned to a selection group and 20 to a reception group. Each *S* in the reception group was yoked with an *S* in the selection group, in that he received instances in the same order selected by his yoked counterpart.

The same eight problems were administered, in random order, to all *S*s in the selection group. The *S*s in the reception group received the problems in the same order as their yoked counterparts in the selection group. The solution for each problem was one value of one dimension, e.g., "red," or "square," or "cross." At the beginning of each problem, *S*s were shown either a positive or negative instance of the concept. Of the eight problems, four were negative examples and four were positive examples. These negative and positive examples were presented in a random order to each pair of *S*s.

All *S*s were told what attributes defined a concept. The *S*s in the selection group were told to sort the blocks into two columns: one containing all instances of the concept and the other containing all instances that were not the concept. They were also instructed to move only one block at a time. For *S*s in the reception group, the instances were presented successively in the same order as their yoked counterpart had selected them. All *S*s were informed if they made an incorrect classification and were instructed to correct their error.

---

[1] Requests for reprints should be sent to Frank S. Murray, Department of Psychology, Randolph-Macon Woman's College, Lynchburg, Virginia 24504.

Results and discussion.—The results of the present experiment showed that concept learning was more efficient with the reception procedure than with the selection procedure. The efficiency of the two procedures was compared by computing the number of errors made by *S*s on all eight problems. The mean number of errors per *S*s in the reception group was 8.95 and for the selection group was 12.50. Analysis of variance for repeated measures was computed. Significantly more errors were made by the selection group than the reception group, $F\ (1,\ 38) = 6.45$, $p < .05$. Statistical significance was also obtained across problems, $F\ (7,\ 266) = 6.64$, $p < .01$. The mean number of errors for Problem 1, 2.6, was the same for both groups, but decreased for both groups on the next seven problems, revealing a learning-to-learn effect. No significant Groups × Problems interaction was obtained, $F\ (7,\ 266) = .48$. A further analysis revealed that the major source of difference between the groups was in the four problems in which negative instances of the concept were presented as examples, $t = 2.89$, $p < .01$ (two-tailed test). There was no significant difference between groups on the four problems with positive-instance examples ($p > .10$, two-tailed test).

To determine the effects of presenting a positive or negative instance within each group, the mean number of errors made on problems in which positive examples were given was compared with the mean number of errors made on problems in which negative examples were given. For the selection group, the mean number of errors on positive example problems was 3.0, and the mean number of errors on negative problems was 9.5. For the reception group, the respective values were 2.4 and 6.6. Significant differences were obtained between positive and negative example problems within both groups ($p > .01$, two-tailed test). There were approximately three times more errors made on negative example problems than on positive example problems. This would be expected from an information analysis of positive and negative instances (Hovland & Weiss, 1953).

One of the difficulties in comparing different procedures in concept learning studies is the use of different measures employed as dependent variables: numbers of correct hypotheses, trial on which correct hypothesis occurred, and trial on which last error occurred. If a real difference exists in the two stimulus presentation procedures, consistent results should occur with different measures. Some of the preceding measures are closely related to each other.

Using this reasoning, another analysis of the data was made with the following measure: trial on which last error occurred. For the four problems in which positive instances of the concepts were presented, the mean values for the reception and selection groups were exactly the same ($\overline{X} = 22.50$). With negative instances, however, the mean values of the two groups differed greatly; the mean trial on which last error occurred for the four problems was 64.00 and 78.25 for reception and selection groups, respectively. A *t* test for matched groups revealed a significant difference between the means ($p < .05$, one-tailed test).

The reason for the difference in efficiency of learning this very simple concept task with the two procedures is difficult to understand. It might be, as suggested by Huttenlocher (1962), that *S*s in the selection group must learn a two-stage task, while *S*s in the reception group were required to learn only one stage. According to Huttenlocher, *S*s in the selection group have to decide on the manipulation necessary for gaining information relevant to the concept and interpret the results of attribute manipulation. The *S*s in the reception group were required to perform only the latter in the present experiment. This would make the task more difficult for *S*s with the selection procedure than with the reception procedure.

Failure to find group differences with positive examples of concepts seemed to be due to an artifactual ceiling effect of the task used in the present experiment. The concept task used was a very simple one for adult *S*s; only two alternative hypotheses were available whenever a positive example of the concept was presented. With such a simple task, few errors were expected. This was borne out in the experiment; mean numbers of errors for *S*s for the four positive problems were 3.00 and 2.40 for selection and reception groups, respectively. With more difficult tasks, one would expect to obtain differences in the two procedures with positive as well as negative examples of the concepts.

REFERENCES

Bruner, J. S., Goodnow, J. J., & Austin, G. S. *A study of thinking.* New York: Wiley, 1956.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology*, 1953, **45**, 175–182.

Hunt, E. Selection and reception conditions in grammar and concept learning. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 211–215.

Huttenlocher, J. Effects of manipulation of attributes on efficiency of concept formation. *Psychological Reports*, 1962, **10**, 503–509.

(Received April 1, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 573–574

CONCEPT ATTAINMENT AS A FUNCTION OF INSTANCE CONTIGUITY AND NUMBER OF IRRELEVANT DIMENSIONS[1]

ROGER L. DOMINOWSKI[2]
University of Illinois, Chicago Circle

Different sequences of presentation were compared with a task requiring the identification of four categories formed by the conjunctive combination of two values from each of two dimensions. Under conditions in which the number of stimulus presentations needed to define the classification was controlled and responding cues were eliminated, most efficient performance occurred with a high-contiguity sequence in which stimuli from the same category occurred successively. The advantage of high contiguity remained constant as the number of irrelevant dimensions varied from three to five. Differences among sequences are attributed to the less ambiguous information provided by a succession of stimuli from the same category, compared to that provided when successive stimuli are from different categories.

Previous research has shown that the rate of concept identification is directly related to the contiguity with which stimuli from the same category are presented (Dominowski, 1965). These findings have typically been interpreted as indicating that identifying the common attributes of category members is more difficult when instances of the same category are separated by instances of other categories because of increased forgetting of relevant stimulus information. The present research is concerned with possible alternative explanations based on methodological considerations.

With the exception of a study by Kurtz and Hovland (1956), previous studies have used a procedure in which the order of stimulus presentation is confounded with testing order. That is, *S* was required to categorize each stimulus as it was presented, then received feedback; as contiguity increases, runs of the same category response tend to occur. Thus, *S* could show improved performance with higher contiguity by learning to give runs of a particular response as well as by remembering more stimulus information. In the Kurtz and Hovland study, *S*s were asked to name the categories and to categorize test stimuli following stimulus sequences varying in contiguity; no differential responding cues were present. Contiguity had a nonsignificant effect on test performance and only a marginally significant effect when verbal descriptions were scored for number of correctly identified dimensions. Thus, the evidence for contiguity effects in the absence of possible differential responding cues is not strong.

Previous research is also unclear with respect to the number of stimulus presentations required to provide sufficient information to define the classification for sequences varying in contiguity. A high-contiguity sequence contains a high proportion of very informative transitions (only irrelevant attributes change from one stimulus to the next), while a low-contiguity sequence is likely to contain less informative transitions (both relevant and irrelevant attributes change); consequently, more stimuli might be required to define the classification as contiguity decreases (Anderson & Guthrie, 1966). The present study was designed to determine the effects of instance contiguity when responding cues and the number of stimuli required to define the classification were controlled. Number of irrelevant dimensions was included as an orthogonal variable, both to examine the effects of this variable under the present conditions and to determine if instance contiguity and number of irrelevant dimensions produce interactive effects.

Method.—The stimuli were geometric forms varying on from five to seven binary dimensions. The relevant dimensions were shape and interior, yielding the conjunctive categories open squares, filled squares, open circles, and filled circles. With three irrelevant dimensions (IDs), number, position, and color were irrelevant; size was added for four IDs, and dot–no-dot was added for five IDs. The forms were drawn on white index cards. For a practice task, 16 cards were contructed containing lines varying in form, orientation, thickness, and arrowhead (two values each).

For each level of ID, three stimulus sequences were created. In each sequence, each block of 16 stimuli (trials) contained 4 stimuli from each category; these 4 stimuli differed such that the category (e.g., open squares) was unambiguously defined. In Sequence 0, the 4 stimuli from a category occurred in succession (0 intervening stimuli); in Sequence 3, stimuli from the 4 categories alternated (e.g., ABCDABCD, etc.; 3 intervening stimuli). Sequence R was constructed by randomizing the 16 stimuli in a block, subject to the restriction that 2 stimuli from the same category could not occur successively (a mean of 2.6 intervening stimuli). For all levels of ID, sufficient information to identify the classification had been given after presentation of 4 stimuli in Sequence 0, after 6 stimuli in Sequences 3 and R.

---

[1] This research was supported by a grant from the University of Illinois Research Board. Robert Horwitz assisted in data collection and analysis. The results were reported at the meeting of the American Psychological Association, San Francisco, September 1968.

[2] Requests for reprints should be sent to Roger L. Dominowski, Department of Psychology, University of Illinois, Chicago Circle, Box 4348, Chicago, Illinois 60680.

There were eight introductory psychology students in each of the nine groups formed by factorially combining three levels of ID with three types of sequences. Each *S* was individually run; *S*s were assigned to groups in alternation as they appeared at the laboratory. Five *S*s were replaced because of failures to follow instructions, and two additional *S*s were randomly eliminated to create equal-sized groups.

The practice task was the same for all *S*s. Following a description of the stimuli (lines), *S* was required to describe test stimuli. After an explanation of the method of forming categories, *S* was asked to state what the categories would be if certain dimensions were selected. Finally, *E* showed the cards, first holding each card above one of four boxes for 10 sec., then placing it in the box. A new stimulus was shown every 15 sec., and cards were not visible in the boxes. Instructions indicated that *S* should tell *E* whenever *S* felt he knew any of the categories, and that stimuli would be shown until *S* had correctly named all four categories (*S* was not required to specify which box represented each category). Each time *S* offered an hypothesis, *E* responded "right" or "wrong." After a brief rest, the same procedure was used for the experimental task.

Results and discussion.—For each *S*, the number of trials to criterion (naming the four categories) was determined. Contiguity had a significant effect, with means of 10.62, 22.33, and 24.12 for Sequences 0, 3, and R, respectively, $F\ (2, 63) = 9.59$, $p < .001$. A Newman-Keuls analysis indicated that Sequence 0 differed significantly from Sequences 3 and R, which did not differ from each other. For three, four, and five IDs, the means were 14.38, 21.42, and 21.29, respectively; although the overall effect of ID was not significant, $F\ (2, 63) = 2.87$, $p < .10$, the linear component was significant, $F\ (1, 63) = 4.14$, $p < .05$. The Sequences × ID interaction was not significant, $F < 1$.

Secondary analyses related *S*s' verbalization of hypotheses to trials to criterion. The percentage of trials on which *S* verbalized was not systematically related to trials to criterion. In six of the nine groups, there was a tendency for *S*s who required more trials to begin verbalizing later. This relationship was most consistent for *S*s shown Sequence R, with correlations of .56, .43, and .58 for 3, 4, and 5 IDs. Early hypothesizing may be an indicator of an active, attribute-testing approach to concept identification, with the results indicating that *S*s who adopt such an approach will be better able to solve the problem. A similar description was offered by Byers and Davidson (1967) in accounting for their finding that requiring *S*s to hypothesize on every trial facilitated concept identification.

The present results indicate that high contiguity results in faster concept identification when possible differential responding cues are eliminated and when the number of stimulus presentations required to define the classification is controlled. Furthermore, the advantage of high contiguity remains constant as the general difficulty of the task increases due to an increase in the number of irrelevant dimensions. Two comments seem in order concerning contiguity effects.

First, neither the present results nor previous research demonstrates that the degree of separation of instances of the same category is an effective variable. In the present study, Sequences 3 and R resulted in equivalent performance, but were quite different types of sequences. In Sequence 3, there were always three intervening stimuli between appearances of stimuli from the same category, whereas in Sequence R this separation was sometimes only one intervening stimulus, sometimes more, averaging 2.6 stimuli. While one might expect that solution would be facilitated by the occurrence of some minimal separations in Sequence R, performance was numerically poorest in this condition. Successful manipulations of instance contiguity have involved variation in the probability of successive occurrence of stimuli from the same category, with difficulty of solution inversely related to this probability (Bourne & Jennings, 1963).

Anderson and Guthrie (1966) offered an explanation for the superiority of a transition in which successive stimuli belong to the same category. When successive stimuli are from the same category, the information provided by the transition is unambiguous—all changed attributes are irrelevant. However, when successive stimuli belong to different categories *and several attributes change*, some unknown proportion of the changed attributes is relevant to the category change. This account suggests that only memory for the previous stimulus is relevant to the effects of manipulating instance contiguity. New data are required to demonstrate that degree of separation (beyond one) is also an effective variable.

REFERENCES

Anderson, R. C., & Guthrie, J. T. Effects of some sequential manipulations of relevant and irrelevant stimulus dimensions on concept learning. *Journal of Experimental Psychology*, 1966, **72**, 501–504.

Bourne, L. E., Jr., & Jennings, P. The relationship between response contiguity and classification learning. *Journal of General Psychology*, 1963, **69**, 335–338.

Byers, J. L., & Davidson, R. E. The role of hypothesizing in the facilitation of concept attainment. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 595–600.

Dominowski, R. L. Role of memory in concept learning. *Psychological Bulletin*, 1965, **63**, 271–280.

Kurtz, K. H., & Hovland, C. I. Concept learning with different sequences of instances. *Journal of Experimental Psychology*, 1956, **51**, 239–243.

(Received April 3, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 575–577

CATEGORY CLUSTERING FOR IMMEDIATE AND DELAYED RECALL AS A FUNCTION OF RECALL CUE INFORMATION AND RESPONSE DOMINANCE VARIABILITY[1]

ROBERT L. HUDSON[2]
Mississippi State University

Words from the Underwood and Richardson norms were presented for free recall. The *S*s were either instructed before word presentation (information-before) or after word presentation (information-after) that the words could be categorized. There was also a no-information control group. The *S*s were presented one of two lists equated for mean response dominance, but differing in response dominance variability. The primary results of interest were that the information-before group clustered significantly more than either the information-after group or the no-information group, between which there was no difference. The information-after group recalled fewer words than either the information-before group or the no-information group, between which there was no difference. The groups receiving the low-variability lists clustered significantly more than the groups receiving the high-variability lists, while words recalled did not differ. These results held for both an immediate recall and a 48-hr. recall. Clustering did not decrease significantly from the immediate to the delayed recall.

Hudson (1968) presented a method for manipulating the categorical relations in a set of words while keeping the words, and hence the level of interitem association, constant. The present study is an extension of this research and was designed to test hypotheses derived from the work of Tulving and Osler (1968) and Freedman and Mednick (1958).

Tulving and Osler (1968) hypothesized that the retrieval cue for a to-be-remembered (TBR) word must be stored with the word at input if it is, in fact, to be a retrieval cue. Their data support this hypothesis for number of words recalled. Significant increases were found when retrieval cues were present at both input and output. If, in the clustering situation, the category name serves as a repetitive retrieval cue, then a group of *S*s given the category names between word presentation and free recall (hereafter called the information-after group) should cluster no more than a no-information group, because the retrieval cues were not present at input.

However, if subsequent opportunities are given for recall of the words, the clustering of the information-after group might be expected to move toward that of the information-before group because a recall serves in a limited way as another input at which time the retrieval cues would be present. All *S*s of the present experiment were given a recall test 20 sec. after presentation (Trial 1) and approximately 48 hr. after presentation (Trial 3). In addition, half of the *S*s were given an intermediate recall trial approximately 25 min. after presentation (Trial 2).

The second hypothesis comes from Freedman and Mednick (1958). They demonstrated that higher response dominance variability led to faster concept acquisition when mean response dominance was equated. They suggest that this is because a word very high in response dominance leads to an earlier correct concept name response. For the category clustering situation, this suggests a significant interaction between level of information and level of response dominance variability of the list. The information-before group should cluster relatively more on the low-response-dominance list, while the information-after and no-information groups should cluster relatively more on the high-response-dominance list.

Method

Subjects.—The *S*s were 96 members of a course called Introduction to Graduate Studies taught the summer of 1968 at Central Washington State College. The median age was 27 yr., and the interquartile range was 25–34 yr. Twelve *S*s were dropped from the analyses either because they failed to follow instructions or by *E* (at random) in order to equate the number per cell.

Materials.—The words were selected from the scaled norms of Underwood and Richardson (1956). Two lists of eight words each were selected from each of the categories round, white, smelly, and small. The two lists of each category were selected so as to differ as much as possible in variability around the mean response dominance value while keeping the two means approximately equal. An attempt was also made to equate the mean number of letters, mean Thorndike-Lorge (1944) frequency, and mean number of competing category names. By competing category is meant the number of names other than the one under consideration that were elicited by the word in question. The words used follow:

---

[1] A part of the data was reported in a paper presented at the meeting of the Southeastern Psychological Association, New Orleans, March 1969.

[2] Requests for reprints should be sent to Robert L. Hudson, Department of Psychology, Mississippi State University, State College, Mississippi 39762.

High variability. *Round:* globe, spool, knob, button, balloon, belly, platter, pot. *White:* milk, bandage, ivory, linen, rice, collar, bread, paste. *Smelly:* ammonia, skunk, sewer, gasoline, sulphur, cigar, zoo, cheese. *Small:* atom, germ, village, minnow, capsule, snail, pollen, kitten.

Low variability. *Round:* dome, knob, head, button, saucer, balloon, belly, platter. *White:* snow, ivory, napkin, linen, rice, collar, bread, frost. *Smelly:* ether, sewer, garlic, gasoline, onion, sulphur, pine, cigar. *Small:* germ, village, minnow, mouse, capsule, pup, pony, pill.

Procedure.—The *S*s were run at the time of their regular class meeting. Four classes were used, two of which were assigned randomly to the high-variability list and two to the low-variability list. Random assignment to information groups was achieved within each class by varying the instructions included in a booklet each *S* was handed. The critical instruction simply stated that the words could be put into categories and gave the category names. Where critical instructions appeared for one group, filler instructions appeared for other groups. The actual procedure of word presentation and recall was the same for all *S*s.

The 32 words of a list were presented by a Kodak Carousel projector four times in four different random orders. The interstimulus interval was 5 sec., of which approximately 4.6 sec. was actual presentation time. The *S*s were allowed 4 min. for recall and were instructed to write the words in a single column in whatever order they occurred in memory. After approximately 20-min. work on a multiple-choice examination, half of the *S*s (two classes) were asked once again to recall the words. The other *S*s were dismissed. At their next class meeting (approximately 48 hr. later), all *S*s were once again asked to recall the words. Four minutes were also allowed for these recall trials.

Results

The clustering index is a standard score indicating in standard deviation units the departure of the observed number of repetitions from that expected by chance (Dunn, in press). It is similar for most recall protocols to the index of Hudson and Dunn (1969).

These clustering data were subjected to a $3 \times 2 \times 2 \times 2$ analysis of variance of the following variables and levels. The information condition consisted of three levels: information-before, information-after, and no-information. The variability condition consisted of two levels: high and low. The interpolation condition (recall after 25 min.) consisted of two levels: present and absent. The trials condition consisted of two levels: Trial 1 (recall after 20 sec.) and Trial 3 (recall after 48 hr.). This variable, of course, was of the repeated measures variety and was so treated. The mean clustering indexes and mean number of words recalled of the various subgroups are presented in Table 1. The analysis indicated that the main effect for information was significant, F $(2, 72) = 5.34$, $p < .01$. The information-before group clustered significantly more than either the information-after group or the no-information group, between which there was no difference. The analysis also indicated a significant main effect for variability, F $(1, 72) = 6.86$, $p < .025$. The low-variability group clustered more than the high-variability group. The remaining main effects and all of the interaction effects were not significant.

An analysis was also conducted using the same conditions on the number of words recalled. The analysis indicated that the main effect for information was significant, F $(2, 72) = 3.47$, $p < .05$. The information-after group recalled significantly fewer words than either the information-before group or the no-information group, between which there was no difference. The main effect for trials was also significant, F $(1, 72) = 66.46$, $p < .005$. Significantly fewer words were recalled

TABLE 1

MARGINAL MEANS AND n'S FOR THE CLUSTERING INDEX (CI) AND NUMBER OF WORDS (NW) RECALLED

Item	Variability condition		Information condition			Interpolation condition	
	High	Low	Before	After	No	Present	Absent
Trial 1							
n	42	42	28	28	28	42	42
CI	1.00	1.77	2.04	1.06	1.07	1.30	1.47
NW	18.17	18.43	19.50	15.93	19.46	18.38	18.21
Trial 2							
n	21	21	14	14	14	42	—
CI	1.12	1.56	2.21	1.18	.63	1.34	—
NW	19.14	19.43	21.07	17.21	19.57	19.29	—
Trial 3							
n	42	42	28	28	28	42	42
CI	.81	1.47	1.73	.68	1.02	1.33	.96
NW	15.12	15.26	16.00	13.14	16.43	15.93	14.45

after 48 hr. than were recalled after 20 sec. The Interpolation × Trials interaction approached significance, $F\ (1, 72) = 2.95$, $p < .10$, which would be expected since interpolation has no meaning until the third trial. None of the remaining main effects or interaction effects were significant or approached significance.

It should be noted that preliminary analysis of both the clustering index and the words recalled on Trial 1 showed no significant differences between the groups that had been treated alike.

The clustering index and the number of words recalled are not completely mathematically independent because the number of words recalled places an upper and a lower limit on the clustering index. Nevertheless, for a given number of words recalled, the clustering index is free to vary over a considerable range. Spearman rank correlation coefficients were computed between the clustering index and the number of words recalled for each of the 24 subgroups on which the analyses of variance were run. These ranged from a −.51 and a −.06 up to .93 and .89. The mean r computed using Fisher's Z transformation was .55. Thus there is a tendency within the subgroups for *S*s who cluster more to recall more words. This is particularly interesting in view of the fact that the information-before group clustered more than the no-information group, yet recalled the same number of words. It is somewhat paradoxical that there is a relationship between organization and recall, but at the same time "artifically" increasing organization does not increase recall.

Discussion

The first hypothesis was supported. The *S*s receiving the information after word presentation apparently did not or could not use this information in organizing their recall. This is not to say that *S*s did not have the TBR words organized into categories, but rather that they apparently were not organized into *E*-defined categories, and hence the category name could not serve as a retrieval cue. Thus it would seem that the organization takes place at the time of stimulus input, and the same recall cue information presented after stimulus input is not effective in changing the organizational aspects of the output of *S*. It should be noted that none of the interaction effects of the analysis on the clustering data were significant, which indicates that the information-after group does not move significantly toward the information-before group from Trials 1 to 3. This might still be expected to occur if there were another input preceding the later recall.

Slamecka (1968) has proposed another, quite different, model to account for the organization often observed in free recall studies. According to this model, organization is not achieved at storage, but rather at retrieval as a result of the search plan. Slamecka (1968) has further suggested, while discussing incidental learning, that if an *S* were given an efficient retrieval plan at the time of recall, his number of words recalled might be expected to increase. The present study was not designed to provide a differential test of the two types of theory. One might suppose, however, that the information-after group of the present study could be considered to be a group given a retrieval plan at the time of recall. The fact that this information-after group evidenced little *E*-defined organization could be interpreted as evidence that a good search plan is not enough, with the implication that words have to be organized at input if organization is to be evidenced at recall. On the other hand, however, one might suppose that *S*s ignore the search plan presented at recall (empirically, the category names) in favor of the search plan they formed during stimulus input. Thus, one would still expect little or no *E*-defined organization to be present. There is evidence, however, indicating that the information-after group did not simply ignore the presentation of the category names. This evidence is that the information-after group showed a significant decrease in number of words recalled. It is also interesting to note that this decreased recall held for both Trials 1 and 3 and was not affected by the presence or absence of an interpolated Trial 2. Tulving and Osler (1968) also found decreased recall for a group given recall cues only at the time of recall, but offered no explanation for the inhibitive action of the recall cues.

The second hypothesis of the present study was not supported. The *S*s receiving the low-variability words clustered significantly more than *S*s receiving the high-variability words, and there were no significant interactions. The reason for this is not immediately apparent, and the degree of difference is particularly striking when one considers that there is approximately a 50% overlap between the two lists of words. It should also be noted that for number of words recalled, there was no difference between the two lists, which indicates that the matching of the two lists on the relevant variables was successful.

REFERENCES

Dunn, J. E. A compound multiple runs distribution. *Journal of the American Statistical Association*, in press.

Freedman, J. L., & Mednick, S. A. Ease of attainment of concepts as a function of response dominance variance. *Journal of Experimental Psychology*, 1958, **55**, 463–466.

Hudson, R. L. Category clustering as a function of level of information and number of stimulus presentations. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 1106–1108.

Hudson, R. L., & Dunn, J. E. A major modification of the Bousfield (1966) measure of category clustering. *Behavior Research Methods and Instrumentation*, 1969, **1**, 110–111.

Slamecka, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, **76**, 504–513.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

Tulving, E., & Osler, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 1968, **77**, 593–601.

Underwood, B. J., & Richardson, J. Some verbal materials for the study of concept formation. *Psychological Bulletin*, 1956, **53**, 84–95.

(Received April 21, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 578–580

NUMBER OF FOOD PELLETS AND THE PARTIAL REINFORCEMENT EXTINCTION EFFECT AFTER EXTENDED ACQUISITION[1]

ABRAM AMSEL,[2] C. THOMAS SURRIDGE, AND JAMES J. HUG

University of Toronto

As a follow-up to an earlier experiment, six groups were rewarded partially or continuously with 1 gm. of food, given in the form of either 1, 10, or 24 pellets. In this experiment, unlike the earlier five-trial experiment, training was over 80 trials at 1 trial a day, and the groups did not differ in resistance to extinction, the partial reinforcement extinction (PRE) effect appearing in all number-of-pellet conditions.

It has recently been shown by the present investigators that in partial reinforcement experiments involving only five acquisition trials, the number of food pellets making up a 1-gm. reward is a critical factor in determining whether or not a partial reinforcement extinction effect (PRE) will appear (Amsel, Hug, & Surridge, 1968). It was argued that when rats are offered multiple food pellet rewards (e.g., Padilla, 1967) or wet mash rewards (e.g., McCain, 1966), they tend to make several approaches (back-and-forth movements) to the food cup on a single trial, so that functionally the few trials become many. This might provide at least a partial explanation for the occurrence of the PRE under some few-trial partial reinforcement conditions. Assuming that number of pellets affects the rate of conditioning, but not the asymptotic strength of anticipatory reward (r_R-s_R) and frustration (r_F-s_F), when total magnitude of reward has been equated, differences in resistance to extinction as a function of number of pellets should be attenuated after extended acquisition. Amsel et al. tested this implication by factorial variation of percentage of reward (50% or 100%) and number of pellets (one 1,000-mg., ten 97-mg., or twenty-four 45-mg.) over 80 acquisition trials.

Method.—Sixty naive, male albino rats, 80–90 days old at the beginning of the experiment, were randomly assigned to six groups of 10 each. The 3-ft. black runway and the experimental procedure were as described earlier (Amsel et al., 1968), with the exception that there were 80 acquisition and 32 extinction trials in the present case. All 50% conditions involved a recursive R R N N R N N R N R R N N N R R sequence, but different *S*s in the same group began the cycle at different points, so that 40–60% of partial *S*s received either R or N on any day. The *S*s were run in squads, which contained an equal number of *S*s from all groups, of 24, 24, and 12; running order was randomized over days, and the intertrial interval was 24 hr.

Number of approach responses (back-and-forth movements) to the food cup was recorded by placing an infrared photobeam, which activated a counter, directly over the proximal lip of the food trough. The counter registered whenever the photobeam was broken by *S*'s head or some other portion of its body.

Results and discussion.—Separate measures of start, run, and goal time were converted to speeds and were analyzed over the 20 blocks of four trials. During acquisition, partial groups ran faster in the start and run segments, $F\ (1, 54) = 5.13$ and 6.84, both p's $< .05$. Response speeds did not vary as a function of number of pellets (all p's $> .10$), although the Percentage Reward × Number of Pellets interaction was significant in the goal measure, $F\ (2,\ 54) = 3.60,\ p < .05$, which resulted from superiority of partial groups receiving 1 or 10, but not 24, pellets.

Extinction speeds are presented in Fig. 1 in terms of average speed over all three runway segments. Partial groups ran faster in all segments, $F\ (1, 54) = 110.51$, 101.55, and 92.42, for start, run, and goal, respectively, all p's $< .01$. Greater resistance to extinction after partial reinforcement training was also reflected by the Percentage Reward × Blocks interaction, $F\ (7,\ 378) = 7.20$, 9.36, and 17.55, respectively, all p's $< .01$. Unlike the Amsel et al. (1968) few-trials experiments involving the same variables, there was no Percentage of Reward × Number of Pellets interaction (all p's $> .10$).

Goal-approach data are presented in Fig. 2, separately for rewarded and nonrewarded trials, for the 1- and 10-pellet groups. Data for the 24-pellet groups were contaminated by jaw movements when *S* ate over the food cup without moving back and forth, and are not presented. The 10-pellet groups obviously made more approach responses than the 1-pellet groups on rewarded trials, both on the first 5 trials that correspond to the previous study (Amsel et al., 1968) and over the entire 80 acquisition trials. Differences between the 10-pellet groups, which appeared rela-

---

[1] This study was supported by Research Grants GB-3772 from the National Science Foundation and APA-72 from the National Research Council of Canada. The authors wish to acknowledge the assistance of Howard Glazer and Susan Izumi in the collection of the data.

[2] Requests for reprints should be sent to Abram Amsel, who is now at the University of Texas at Austin, Department of Psychology, Austin, Texas 78712.

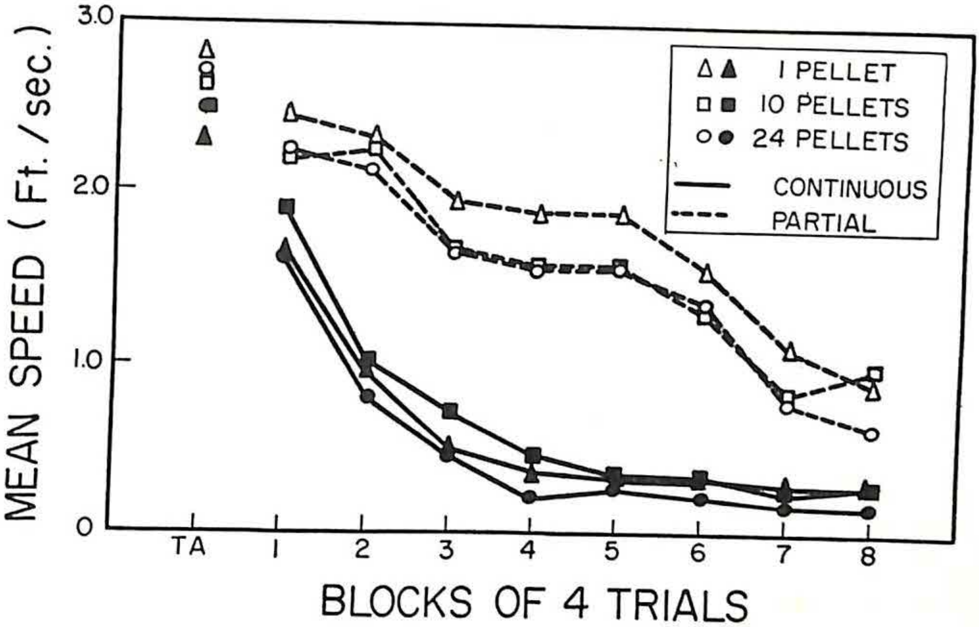


Fig. 1. Total alley speed for terminal acquisition and extinction trials. (TA represents the last four acquisition trials.)

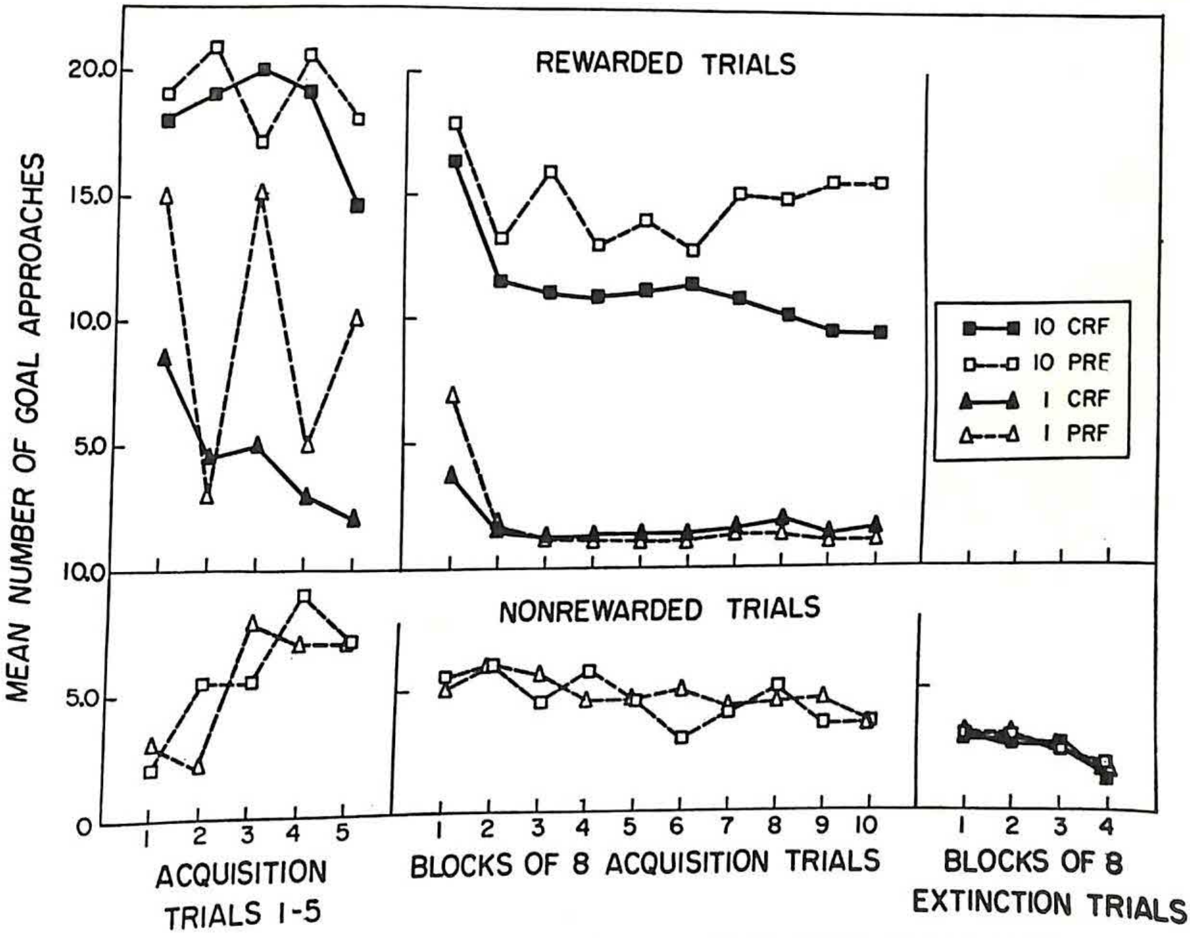


Fig. 2. Number of goal approaches for 1- and 10-pellet groups. (Trial-by-trial data are presented for the first five acquisition trials in the left-hand panels.)

tively substantial toward the end of training, were short of significance, $F\ (1, 18) = 3.26$, $.05 < p < .10$; the Percentage × Blocks interaction for the 10-pellet groups was also nonsignificant, $F\ (9, 162) = 1.06$, $p > .10$. The lack of any systematic differences between the 1- and 10-pellet groups on nonrewarded acquisition trials, or during extinction, is clear in Fig. 2.

While the greater frequency of goal approaches on rewarded trials with multiple-pellet rewards lends some support to the hypothesis that number of food pellets affects the rate of conditioning of $r_R\text{-}s_R$, the lack of a number-of-pellets effect on nonrewarded trials does not necessarily vitiate our argument that multiple-pellet rewards result in counterconditioning of $r_F\text{-}s_F$ to the instrumental response, even when acquisition has been very limited. It remains conceivable that $r_F\text{-}s_F$ is present on very early acquisition trials only when a multiple-pellet reinforcement, or its equivalent, is used, for the reason that even early nonrewarded trials will result in primary frustration (R_F) in the presence of stronger $r_R\text{-}s_R$ conditioned by multiple-pellet rewards.

However, it remains possible that the PRE following a very small number of acquisition trials is entirely dependent on the presence of sufficient r_R to make nonreward frustrating in extinction and that this kind of PRE has nothing to do with the counterconditioning of s_F to approach in partial reinforcement frustration (PRF) training. We are suggesting that in the "PRE" with a small number of trials, the PRF *S*s extinguish more slowly, if at all, because they have developed insufficient r_R by the end of "acquisition" to make nonreward frustrating in "extinction."

The speed data from the present experiment show that the direction and magnitude of the PRE are unaffected by manipulation of number of food pellets after extended acquisition. This result is not surprising, given that total magnitude of reward has been equated across the number-of-pellet conditions. Previously (Amsel et al., 1968), it was pointed out that experiments in which magnitude of reward was defined by number of food pellets could have confounded number of trials with magnitude of reward. The present data indicate that this possible confounding does not affect performance after extended acquisition.

REFERENCES

AMSEL, A., HUG, J. J., & SURRIDGE, C. T. Number of food pellets, goal approaches, and the partial reinforcement effect after minimal acquisition. *Journal of Experimental Psychology*, 1968, **77**, 530–534.

MCCAIN, G. Partial reinforcement effects following a small number of acquisition trials. *Psychonomic Monograph Supplements*, 1966, **1**, 251–270.

PADILLA, A. M. A few acquisition trials: Effects of magnitude and percent reward. *Psychonomic Science*, 1967, **9**, 241–242.

(Received April 28, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 581–583

CHOICE RESPONSE TIMES AS FUNCTIONS OF INTRALIST SIMILARITY, STIMULUS TYPE, AND NUMBER OF EQUALLY PROBABLE ALTERNATIVES[1]

BARRY GHOLSON[2]
State University of New York at Stony Brook

AND

RAYMOND H. HOHLE
University of Iowa

Stimuli drawn from response-defined interval scales of size and lightness dimensions were used to investigate effects of visual similarity, stimulus type, number of equally probable alternatives, and practice on choice response time (RT). Results indicated a greater increase in RT with number of equally probable alternatives for lists containing similar items than for lists containing dissimilar items. There was also a List Length × Practice interaction, indicating that the slope of the function relating RT to list length decreased across daily sessions.

Numerous investigators have noted that while average verbal response time (RT) increases with number of equally probable alternatives, the slope of the regression line, when RT is plotted as a function of number of alternatives, may differ considerably depending on the type of stimuli used. Several investigators (see Smith, 1968) have suggested that under some conditions at least, differential increases in RT with list length may be attributable to differing levels of visual similarity of the stimulus alternatives. Gholson and Hohle (1968) found greater increases in mean RT with increases in list length for stimulus ensembles composed of hues or forms than for sets composed of hue names or form names. It was suggested that individual stimuli within a set of hues (or forms) are highly similar to each other, differing slightly on only one or two dimensions, while individual stimuli within a set of hue names (or form names) are more dissimilar, differing on several dimensions including letters, letter position, length, etc. The assumption here was that RT increases monotonically with interstimulus interference due to visual similarity of the stimulus alternatives and that this increase in RT is reflected in the *slope* of the function relating RT to number of alternatives, rather than in its intercept.

In the present research, stimuli were drawn from response-defined scales that were assumed to provide psychological intervals of stimulus dissimilarity. The equal discriminability (ED) method (see Thurmond & Alluisi, 1963) was used to define interval scales for the lightness (shades of gray) and size (circle diameters) dimensions. Both ED scales were assigned the same unit since the two sets of stimuli were scaled under exactly the same experimental conditions. Stimuli selected from these scales were used to investigate effects of differing levels of stimulus dissimilarity on the slope of the function relating RT to number of equally probable alternatives.

Method.—The *S*s were 48 females (paid volunteers), 15–18 yr. of age. They were randomly assigned to four groups of 12 each. Each *S* was presented three lists of stimuli, each containing a different number of potential alternatives.

Alternative stimuli composing the different lists were chosen on the basis of preliminary scaling to sample high and low levels of dissimilarity on each scaled dimension. Shades of gray were obtained by combining Liquitex Polymer Emulsion Artists' Colors (Permanent Pigments, Cincinnati 12, Ohio) mars black with titanium white. Differing circle diameters were obtained via a calibrated adjustable diaphram (see Table 1).

Stimuli were presented via a three-field tachistoscope. Visible areas of all fields were mars black except as follows: Reaction stimuli and black diaphram leaves were visible in Field 1; and a horizontal white stripe that served as a warning signal was visible in the bottom of Field B. Trials began with a 6.5-sec. presentation of Field 2. Next, Field B containing the white stripe warning signal was switched on for 1.0 sec. After this foreperiod interval, Field 1 of the tachistoscope, containing the reaction stimulus, was switched on and a chronoscope started simultaneously. The *S*'s verbal recognition response then activated a voice key, which terminated the reaction stimulus (i.e., returned the tachistoscope to Field 2) and stopped the clock. The clock reset automatically each trial at onset of Field B.

Each *S* was presented three lists of stimuli, each containing a different number of equally probable alternatives (two, three, and four items). Two groups were assigned lists in which stimuli differed on the size dimension, while lists assigned to the other two groups contained items differing on the lightness dimension. For one of the groups as-

---

[1] This paper, supported in part by United States Public Health Service Research Grant HD-2681-01, is based on part of a thesis submitted by the first author in partial fulfillment of the requirements for the PhD degree at the University of Iowa. The second author directed the thesis. Copies of the thesis may be obtained from the first author upon request.

[2] Requests for reprints should be sent to Barry Gholson, Department of Psychology, State University of New York at Stony Brook, Stony Brook, Long Island, New York 11790.

TABLE 1

ED Scale Values and Physical Values of Stimuli Composing the Four-Item List Assigned to Each of the Four Groups

Dissimilarity groups	ED scale values	Circle diameters (in $\frac{1}{32}$ in.)	Percentage of titanium white[a]
Low	1 3 5 7	14.25 15.60 18.25 20.60	97.8 91.7 75.6 52.9
High	1 4.5 8 11.5	14.25 17.20 21.90 26.20	97.8 83.0 42.2 10.5

[a] By weight.

signed stimuli varying in size and one of the groups assigned stimuli varying in lightness, the alternatives of each list were similar (low dissimilarity), while the remaining two groups were presented lists containing dissimilar alternatives (high dissimilarity).

Stimuli included in the four-item lists assigned to each group of 12 *S*s were an ordered series drawn from the appropriate response-defined scale. Dissimilarity values of stimuli were obtained, following Thurmond and Alluisi (1963), by taking the difference in units of the ED scale (ED units) between appropriate items. For the low-dissimilarity groups, stimuli in the series were selected in steps of 2 ED units, while the series assigned the high-dissimilarity groups were in steps of 3.5 ED units. Specific stimuli included in the four-item list assigned to each group can be seen in Table 1. Numbers in the rows of the table labeled "ED scale values" denote stimuli defined by the ED scaling procedure, with 1 representing the smallest circle (lightest gray) and larger numbers representing increasingly larger circles (darker grays) drawn from the response-defined interval scales.

Stimuli composing the two- and three-item lists presented to *S*s of each group were the 12 combinations of four items two and three at a time, combined under the restriction that each list for a given *S* contain items in the next smaller list. All three lists (i.e., two, three, and four items) were presented in each of three daily sessions. List presentation order within each of two sequences was counterbalanced for each group between *S*s within sessions and within *S*s over days.

A unique verbal recognition response was required upon presentation of each list item. Response syllables were *boo, bun, bev,* and *baw.* Syllables were randomly assigned to stimuli in the four-item list, with consistent assignment in smaller lists such that a given stimulus was assigned the same syllable in each list in which it appeared.

Each *S* came individually to the experimental room and was seated facing into the tachistoscope. On Day 1, *S* was told that during each experimental session, circles of different sizes (gray circles differing in lightness) would appear on the screen in front of her and that she should identify each by calling out an appropriate name. It was emphasized that she should always respond as fast as possible but avoid making errors. The *S* was instructed that at the beginning of each new series of trials, she would see the alternative stimuli (two, three, or four items) that could appear during the following series and would be told the appropriate name for each.

At this point *S* was told how many items would be included in the first list and each was exposed 2 times for 5 sec., with *E* calling out the appropriate response syllable. A warm-up series in which each item in the given list occurred 5 times was then initiated. A 30-sec. rest interval followed, during which *S* was cautioned to respond as fast as possible while avoiding errors. Each reaction stimulus in the given list then appeared 20 times in a quasi-random sequence, with *E* recording RTs in milliseconds and changing stimuli between trials. An identical procedure was followed for each of the three lists in each daily session. All errors were immediately corrected and recorded by *E*, with the trial repeated later in the same block. Thirty-second rest periods followed each block of 40 trials.

Results and discussion.—This section deals with results obtained for the two stimuli common to all three lists for a given *S*. A four-factor analysis of variance was performed on mean RTs derived from each *S*'s 20 RTs per stimulus per session at each level of list length. Means of the two items common to all three lists were averaged for analyses of variance. There were two between-*S*s treatments: stimulus type (lightness *vs.* size dimension) and dissimilarity level (low *vs.* high). List length and daily sessions were within-*S*s treatments. Table 2 contains a summary of this analysis of variance.

The major findings may be summarized as fol-

TABLE 2

Analysis of Variance

Source	*df*	*MS*	*F*
Between *S*s	47	109338.94	
Dissimilarity (Di)	1	454894.00	4.33*
Stimulus type (S)	1	63562.00	.60
Di × S	1	18.00	.00
Error (between)	44	105010.36	
Within *S*s	384	20079.52	
Days (Da)	2	234757.00	39.18**
Da × Di	2	1830.00	.30
Da × S	2	10553.00	1.76
Da × Di × S	2	1062.00	.17
List length (L)	2	2689420.00	359.24**
L × Di	2	66238.00	8.84**
L × S	2	8004.00	1.06
L × Di × S	2	1323.00	.17
Da × L	4	22635.50	10.55**
Da × L × Di	4	3432.50	1.60
Da × L × S	4	2287.50	1.06
Da × L × Di × S	4	1809.00	.84
Error (within)	352	4441.81	
Error 1	88	5991.13	
Error 2	88	7486.27	
Error 3	176	2144.91	
Total	431	29813.15	

* $p < .05$.
** $p < .001$.

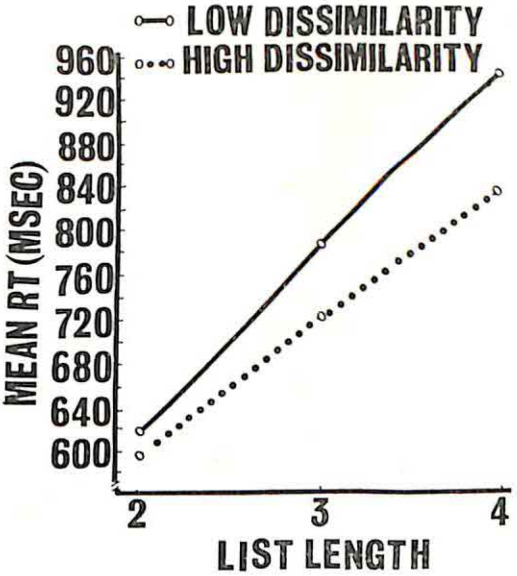


FIG. 1. Mean choice RTs for high and low levels of stimulus dissimilarity as functions of number of equally probable stimulus–response alternatives.

lows: (*a*) the List Length × Dissimilarity Level interaction was significant; (*b*) no effects involving stimulus type were significant; and (*c*) there was a significant List Length × Daily Sessions interaction.

The List Length × Dissimilarity Level interaction (see Fig. 1) indicated that the slope of the function relating mean RT to number of equally probable alternatives was greater for the low- than for the high-dissimilarity level. This result was consistent with the suggestion (Gholson & Hohle, 1968; Smith, 1968) that differential increases in mean RT with list length may be due to differing levels of visual similarity of items included in the stimulus ensemble. Smith (1968) has reviewed data from card-sorting and many:1 mapping tasks that are in accord with those reported here.

The finding of no significant effects involving stimulus type is taken to indicate that at least for the two sets of stimuli used, it was response-defined intralist similarity that accounted for differential increases in mean RT with list length, not stimulus type per se. The generality of this finding may be limited (Smith, 1968), however, since unlike stimuli used in most choice RT experiments, those used here were unidimensional.

The List Length × Daily Sessions interaction (see Fig. 2) indicating that practice reduced the slope of the function relating RT to list length was unexpected (Smith, 1968) and was inconsistent with related research in which 3–10 daily sessions were used (Gholson & Hohle, 1968). In the earlier research, however, required responses were always highly overlearned naming responses (e.g., naming numerals, words, hues, or familiar forms), while in the present study, *S*s were required to associate nonsense syllables with unfamiliar grays and circles. The possibility that practice in associating unfamiliar stimuli and responses may reduce the slope of the RT–list-length function receives some support from a related finding of Davis, Moray, and Treisman (1961). They reported a similar slope reduction in a task requiring imitative responses to acoustically presented unfamiliar nonsense syllables.

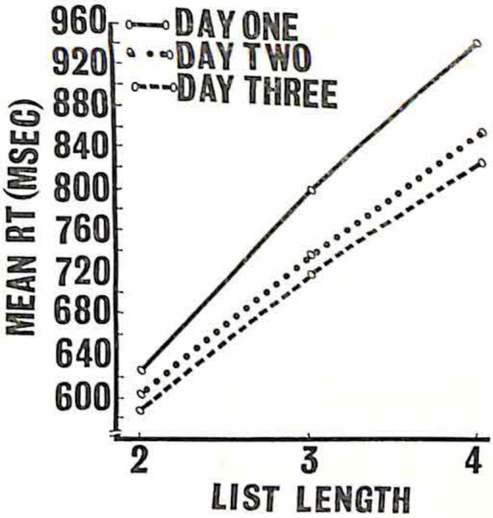


FIG. 2. Mean choice RTs for Days 1, 2, and 3 as functions of number of equally probable stimulus–response alternatives.

One tenable explanation for these findings derives from the acquired distinctiveness of cues (ADC) hypothesis (see Cantor, 1965). It states that learning names for highly similar stimuli tends to lessen generalization of responses (i.e., reduces similarity) among items included in the ensemble. The present findings would follow since, as shown previously (see Fig. 1), similarity is clearly a potent variable affecting the RT-list-length function. The tenability of the ADC hypothesis unfortunately cannot be directly evaluated in the present research since the scaling study was carried out prior to the RT study and used different *S*s. Some indirect support is found, however, in a three-factor analysis of variance performed on error data. There were two between-*S*s factors: stimulus type and dissimilarity level. Daily sessions was a within-*S*s factor. The only significant effect was daily sessions, $F\ (2, 88) = 56.76$, $p < .05$. The error rate was 8.26% on Day 1, 5.15% on Day 2, and 4.01% on Day 3.

REFERENCES

CANTOR, J. H. Transfer of stimulus pretraining in motor paired-associate and discrimination learning tasks. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior,* Vol. II. New York: Academic Press, 1965.

DAVIS, R., MORAY, N., & TREISMAN, A. Imitative responses and the rate of gain of information. *Quarterly Journal of Experimental Psychology,* 1961, **13**, 78–89.

GHOLSON, B., & HOHLE, R. H. Choice reaction times to hues vs. hue names and forms vs. form names. *Perception and Psychophysics,* 1968, **3**, 191–196.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin,* 1968, **69**, 77–110.

THURMOND, J. B., & ALLUISI, E. A. Choice times as a function of stimulus dissimilarity and discriminability. *Canadian Journal of Psychology,* 1963, **17**, 326–337.

(Received April 30, 1969)

Journal of Experimental Psychology
1969, Vol. 82, No. 3, 584–586

EVALUATION OF A MODEL RELATING THORNDIKE-LORGE FREQUENCY AND *m* TO LEARNING[1]

VITO MODIGLIANI
Wesleyan University

AND

ELI SALTZ[2]
Center for the Study of Cognitive Processes, Wayne State University

Forty-two *S*s learned paired-associate (PA) lists in which the stimuli were equated for meaningfulness (m) and varied in frequency (F) of occurrence in the English language. As in a study by Saltz, increased F was accompanied by increased errors during learning ($p < .001$). The present data and those of Saltz were fit to a previously proposed model and compared. Despite the fact that S and R were prior associates in the present study and were unrelated in Saltz, the crucial parametric values in the model remained invariant across the two studies.

Saltz and Modigliani (1967) have proposed a model which related paired-associate (PA) learning to a two-process interpretation of meaningfulness. The first of these processes is concerned with conserving the previous meanings of a word. It is assumed that with greater frequency of occurrence of a word in a given context, the word becomes more resistant to casual new associations. Since a word typically has entered into many previous associations, it is assumed that the tendency to conserve previous meanings is a function of the *average* frequency per previous association, written F/M, where F represents the total frequency of occurrences of a word in the language and M represents the number of different associations to the word. The second process involved in meaningfulness of a word is its differentiation from other words. Differentiation is assumed to be a function of the total frequency of a word, irrespective of context.

The relationship between PA and meaningfulness is then written

$$\text{Errors} = C + a\frac{F}{M} - bF, \qquad [1]$$

where C is a function of the resistance to interference of the new association being learned, while a and b are constants dependent on the procedure used. In the previous studies in this series, F was operationalized as Thorndike-Lorge (1944) frequency, and M as Noble's (1952) meaningfulness (m).

Equation 1 was found to provide a fairly good first approximation to the Saltz and Modigliani (1967) data for response manipulation. However, Equation 1 assumes a linear relationship between errors and F and F/M. Saltz and Modigliani suggested that a logarithmic relationship might provide a better approximation. It is proposed, then, to substitute log $(1 + F)$ and log $(1 + M)$ for F and M in Equation 1 to obtain[3]

$$\text{Errors} = C + a\frac{\log(1+F)}{\log(1+M)} - b\log(1 + F). \qquad [2]$$

The purpose of the present study was to determine the stability of the parameters of Equation 2 across extremely different procedural conditions involving manipulations of *stimulus* meaningfulness. Data reported by Saltz (1967) are compared with data collected in the present study. Stimulus and response terms of the pairs were unrelated in the Saltz study. In the present study, the stimulus and response of each pair were prior associates. Related terms should be learned more rapidly than unrelated. In terms of Equation 2, related terms should affect C, which represents the resistance to interference of the newly learned S-R system. However, the a and b parameters should be unaffected, since they are determined by the tendency to conserve previous meanings of the stimuli and the differentiation of the stimuli, respectively. To prevent too rapid learning, in the present study, presentation rate was increased to 1 sec. (as opposed to 2 sec. in the Saltz study) and number of pairs was increased to 14 (from 10 in the previous study).

Method.—The material consisted of three basic lists of 14 pairs each. The stimuli of each list (taken from the 100 words scaled by Saltz, 1967) had approximately the same m value: m level was 6.00–6.99 for one list, 7.00–7.99 for a second list, and 8.00–8.99 for a third. Within each list there were five levels of stimulus F: 1–4, 30–40, 100–200, 399–750, and 1,000–2,000 occurrences per 4½ million in the Thorndike-Lorge (1944) word count. In each list there were two stimulus words at the lowest F level and three at each of the other F

---

[1] This study was supported by National Science Foundation Grant GD 6584.

[2] Requests for reprints should be sent to Eli Saltz, Department of Psychology, Wayne State University, Detroit, Michigan 48202.

[3] Log $(1 + F)$ is used instead of log (F) because log $(1 + F)$ has the desired property of being always greater than zero for F greater than zero. Similar reasoning applies for using log $(1 + \text{M})$ rather than log (M). The ratio F/M in Equation 1 and the ratio log $(1 + M)$ in Equation 2 are assumed to be always finite.

levels. Each list of stimuli was associated with two different sets of responses, half of the *S*s learning one set of responses, half learning the other. The criterion adopted for the selection of a word as a response to a given stimulus was that at least five, but not more than eight, *S*s had given it as an association to the stimulus in the Saltz study. This criterion insured that the response would be neither one of the most nor one of the least frequently given associations. In three cases the same words (i.e., *coin, number,* and *order*) had to be used in both sets of responses because they were the only ones satisfying the criterion. Examples of S-R pairs are: FILTER-*water*, PROBLEM-*trouble*.

The *S*s, 42 students in an introductory psychology course at Wayne State University, were assigned randomly to conditions. Lists were presented on a Stowe memory drum with a 1-sec. anticipation period (stimulus presented alone) followed by 1-sec. simultaneous presentation of S-R pair. The intertrial interval was 4 sec. The *S*s were run to the criterion of one errorless trial.

Statistical analysis.—The present study involves a 3 × 5 design (3 levels of m, 5 of F), as did that of Saltz and Modigliani (1967) and Exp. I and II of Saltz (1967). In the present paper, all of these experiments were subjected to the identical type of least-square analysis for goodness of fit to Equation 2. The analysis for each experiment involves minimizing the sum of squares for the differences between predicted and obtained errors for each of the 3 × 5 cells. Inserting the values of M and F for each cell (M_h, F_k) into Equation 2, one obtains a predicted value, $\hat{E}_{hk}'$, for the mean errors in that cell. (In computing $\hat{E}_{hk}$, the following average values of the independent variables M_h and F_k were assumed: $M_1 = 6.5$, $M_2 = 7.5$, $M_3 = 8.5$; $F_1 = 2$, $F_2 = 35$, $F_3 = 150$, $F_4 = 550$, $F_5 = 1{,}500$.) The term to be minimized is

$$\sum_{h=1}^{3}\sum_{k=1}^{5}(\hat{E}_{hk} - E_{hk}),$$

where h is a value of M, k is a value of F, $\hat{E}_{hk}$ is the predicted mean errors for the cell hk, and E_{hk} is the obtained mean errors for that cell.

Results.—As in Saltz (1967), the learning data indicated that increased stimulus F led to a greater number of errors at each m level. This effect was highly significant, $F\ (4, 156) = 16.23$. While errors decreased as a function of m, neither this effect nor the $F \times m$ interaction approached significance.

The primary purpose of the present study was to determine the stability of the parameters of Equation 2. Figure 1 shows the mean errors at each F level for each of the three studies. Table 1 presents the values for C, a, and b as obtained for the three studies in the curve fitting procedure de-

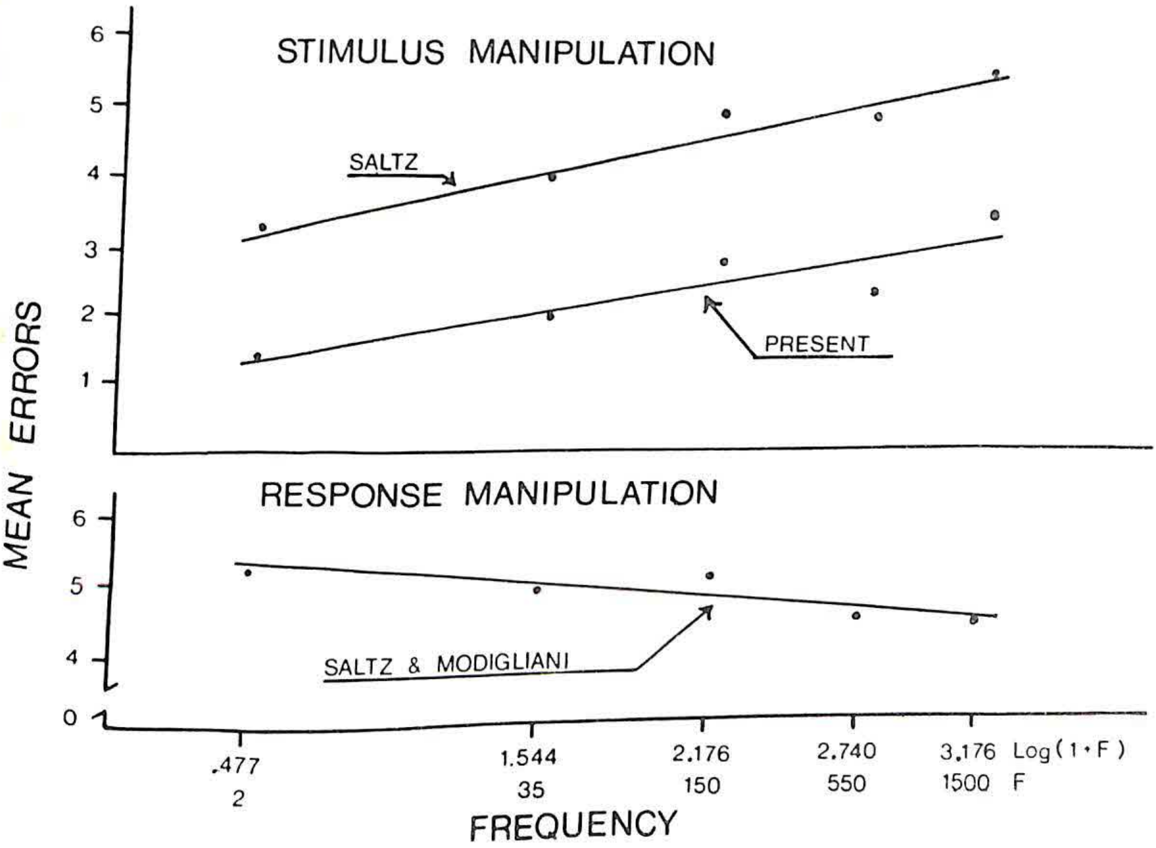


FIG. 1. Mean errors at each F level for each of the three studies.

TABLE 1

Constants for the Logarithmic Fit

Study	Manipulation	Constants		
		C	a	b
Present	Stimulus	1.01	4.01	3.69
Saltz (1967)	Stimulus	2.80	5.03	4.66
Saltz & Modigliani (1967)	Response	5.49	1.63	2.09

scribed earlier. The values of a and b for the stimulus studies (i.e., Saltz, 1967, and present study) were not significantly different. This is shown by a cross validation procedure. The values of a and b from the present study were used to predict Saltz's data. The means thus predicted and the observed means were not significantly different. The F ratio for the goodness-of-fit test was less than 1 for the frequency effects and $F(2, 42) = 1.32$ for the meaningfulness effects. On the other hand, the two C values (which could be compared directly) were significantly different, $F(1, 123) = 7.01$. The values of the constants a and b for the response study were clearly significantly different from those of the stimulus studies; in the former study, $a < b$ produced a significant decrease in errors as F increased, while in the latter, $a > b$ produced a significant increase in errors as F increased.

A comparison of the goodness of fit for the linear vs. logarithmic functions was also undertaken. For each of the three studies (Saltz, 1967; Saltz & Modigliani, 1967; and the present study), the logarithmic form fits better than the linear. For the log frequencies, over 15 points (5 in each experiment), only 1 produced a significant difference between predicted and observed means.

Conclusions.—The data of the present study support those of Saltz (1967) in finding a significant tendency for increased stimulus F to produce increased errors in learning. Further, the a and b parameters of Equation 2 proved to be relatively invariant between the two studies, despite the sizable procedural differences. The difference in difficulty between the studies appears to be attributable to the C parameter. The fact that the same equation can be used to describe either the stimulus or the response manipulation also lends support to the general formulation. It supports the positing of two different processes, conservation of previous meanings and differentiation, working in competition with one another but having strengths which are dependent on the locus of list manipulation.

REFERENCES

Noble, C. E. An analysis of meaning. *Psychological Review*, 1952, **59**, 421–430.

Saltz, E. Thorndike-Lorge frequency and *m* of stimuli as separate factors in paired-associates learning. *Journal of Experimental Psychology*, 1967, **73**, 473–478.

Saltz, E., & Modigliani, V. Response meaningfulness in paired associates: T-L frequency, *m*, and number of meanings (*dm*). *Journal of Experimental Psychology*, 1967, **75**, 313–320.

Thorndike, E. L., & Lorge, I. *The teachers' word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received May 6, 1969)